Bioethanol Production

Characterisation of a Bifunctional Alcohol Dehydrogenase from Geobacillus thermoglucosidasius

Extance, Jonathan

Award date:
2012

Awarding institution:
University of Bath

Bioethanol Production: Characterisation of a Bifunctional Alcohol Dehydrogenase from *Geobacillus thermoglucosidasius*

Jonathan Peter Extance

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

September 2012

Signed by the author

ACKNOWLEDGEMENTS

I would like to acknowledge the people who have contributed to the completion of this project. My fantastic supervisors Prof Michael Danson and Dr David Hough, I am grateful to you both for the opportunity to carry out a PhD project in your lab; the generous amount of time, advice and patience you have both shown me is humbling. I am hugely grateful for the help and support of Dr Susan Crennell during the structural work carried out in this project, particularly for assistance with the dark art of crystallography and for helpful discussions. Financial contributions from the BBSRC and TMO Renewables Ltd are also recognised.

Special thanks are extended to all in Lab 1.33 both past and present, particularly Dalal Binjawhar, Giannina Espina Silva, Dr Tracey Goult, Chris Hills, Nia Marrott, Dr Philippe Mozzanega, Dr Charlie Nunn, Dr Karl Payne, Dr Sylvain Royer, Chris Vennard, Dr Natasha Voina and Carolyn Williamson. The lab has always been a pleasure to work in and all of your suggestions, assistance, encouragements, smiles, and jokes have made this project achievable. Project students Emily Rowe and Melanie Scarisbrick also worked on the project. Prof David Leak and Dr Jeremy Bartosiak-Jentys' assistance with *Geobacillus* transformation work and the donation of expression constructs is also appreciated.

I am thankful for the kind support and guidance from the whole team at TMO Renewables, especially the strain development, fermentation and analytical teams. Particular recognition is given to Dr Gareth Cooper, Dr Kirsten Covington, Dr Roger Cripps, Dr Steve Martin, Dr Jo Neary and Dr Alex Pudney, all of whom have provided excellent assistance during this project. The foresight, wisdom and motivation provided by the late Prof Tony Atkinson is also to be recognised.

I would also like to thank God, all my friends and family who have all been an incredible support through what has been a rewarding but at times difficult journey. Mum, Dad, Chris and Angie you are the best parents (and in-laws) anyone could ever ask for, thanks for always being there for me to rant at or to ask your advice.

Lastly, my wonderful wife Lizzie, you have always encouraged and loved me, you have dragged me through tough times and laughed with me in good times. I will be eternally grateful for everything you've done for me, thank you!

CONTENTS

Acknowledgments i
Contents ii
Abstract vii
List of abbreviations viii

Chapter 1
Introduction

1.1 Energy demands 1
1.2 Biofuels 2
1.3 Second-generation bioethanol 4
1.4 *Geobacillus thermoglucosidasius* 6
1.5 ADHE 8
1.5.1 Sequence analysis 10
1.5.2 Substrate specificity 15
1.5.3 Relative activities of the two domains 15
1.5.4 Divalent metal ions 16
1.5.5 Regulation of expression 16
1.5.6 Spirosome formation 18
1.5.7 Structural studies 19
1.5.8 Pyruvate formate lyase inactivation 19
1.5.9 Ethanol tolerance 19
1.6 Project aims and objectives 20

Chapter 2
General materials and methods

2.1 General laboratory reagents 21
2.2 Strains, media and culture growth 21
2.2.1 *Escherichia coli* (*E. coli*) 21
2.2.2 *G. thermoglucosidasius* 21
2.2.3 Cloning strains 22
2.2.4 Bacterial growth media 22
2.2.5 Strain storage 23
2.3 Molecular biology 24
2.3.1 Primer design 24
2.3.2 Polymerase chain reaction (PCR) 24
2.3.3 Site-directed mutagenesis 26
2.3.4 Agarose gel electrophoresis 26
2.3.5 DNA purification 26
2.3.6 A-tailing 26
2.3.7 pGEM®-T easy 27
2.3.8 Ligation 27
2.3.9 Ethanol precipitation 27
2.3.10 Blue/white colony screening 28
2.3.11 Plasmid DNA preparation 28
2.3.12 Restriction digests 28
2.3.13 DNA sequencing 29
2.3.14 Preparation of electrocompetent *E. coli* 29
2.3.15 Electroporation (transformation) of electrocompetent *E. coli* strains 29
2.3.16 Transformation of chemically-competent *E. coli* strains 29
2.3.17 Preparation of electrocompetent *Geobacillus sp* 30

2.3.18 Electroporation (transformation) of electrocompetent *Geobacillus* strains 30
2.4 Protein expression 30
2.5 *Geobacillus* culturing techniques 31
2.5.1 Tube fermentation method 31
2.5.2 Overnight shake-flask method 31
2.6 Cell extract preparation 32
2.7 Sodium Dodecyl Sulphate PolyAcrylamide Gel Electrophoresis (SDS-PAGE) 32
2.8 Estimation of protein concentration 33
2.9 Metal-affinity purification 34
2.10 Dialysis 34
2.11 Ion-exchange chromatography 35
2.12 Gel filtration 35
2.13 Enzyme activity assays 35
2.13.1 Acetaldehyde dehydrogenase (aldDH) 36
2.13.2 Alcohol dehydrogenase (ADH) (Acetaldehyde reduction) 37
2.13.3 Acyl-CoA preparation and quantification 38
2.13.4 NAD^+ : CoA released ratio assays 38
2.13.5 Optimum temperature (T_{opt}) assays 39
2.13.6 Thermostability (T_{stab}) assays 39
2.13.7 Kinetic parameter determination 40

Chapter 3

Characterisation of the ADHE enzyme from *Geobacillus thermoglucosidasius*

3.1 Introduction 42
3.2 Materials and methods 43
3.2.1 TM242 fermenter growth protocol 43
3.2.2 Protein expression 44
3.2.3 Mass spectrometry analysis of SDS-PAGE gel excised protein bands 44
3.2.4 Dynamic light scattering 45
3.2.5 NanoSight analysis 45
3.2.6 Transmission electron microscopy (TEM) 46
3.3 Native ADHE 47
3.3.1 Results 47
3.3.2 Assay development 47
3.3.3 Optimum temperature of native ADHE activity 48
3.3.4 Tube fermentation analysis 49
3.3.5 SDS-PAGE analysis 49
3.3.6 Kinetics of ADHE in cell extracts 50
3.3.7 Preliminary substrate specificity experiments 56
3.3.8 Purification of native ADHE 57
3.3.9 Estimation of the Mr of the ADHE protein 60
3.3.10 Kinetics of partially-purified ADHE 60
3.3.11 TM400 enzyme assays 64
3.3.12 Discussion 64
3.4 Recombinant ADHE 67
3.4.1 Results 67
3.4.2 Gene cloning and protein expression trials 67
3.4.3 Protein purification 68
3.4.4 Mass spectrometry (MS) of co-eluting bands 69

3.4.5 Kinetics of partially-purified recombinant ADHE 70
3.4.6 Optimum temperature of recombinant ADHE activity 74
3.4.7 Zinc stimulation of ADH activity 74
3.4.8 Discussion 75
3.5 Investigation into the multimeric assembly of ADHE proteins 77
3.5.1 Results 77
3.5.2 Gel filtration 77
3.5.3 Dynamic light scattering (DLS) 79
3.5.4 Nanosight analysis 80
3.5.5 Electron Microscopy 81
3.5.6 Dissociation experiments 82
3.6 Discussion 83
3.7 ADHE general discussion 84

Chapter 4

Isolation and characterisation of independent ADHE domains

4.1 Introduction 89
4.2 Materials and methods 91
4.2.1 ADHE fragment cloning 91
4.2.2 Aerobic *Geobacillus* expression 92
4.2.3 Micro-aerophilic *Geobacillus* expression 92
4.2.4 Partial purification and sizing of active ADH fragments 92
4.2.5 His-tagged fragment cloning 93
4.3 *Geobacillus* expression hosts results 94
4.3.1 Fragment cloning 94
4.3.2 Fragment expression trials (*Geobacillus*) 95
4.3.3 Kinetic characterisation of the ADH domain 99
4.3.4 Protein fragment sizing experiments 102
4.4 *E. coli* expression host results 104
4.4.1 Fragment cloning (pET28a) 105
4.4.2 Recombinant protein expression trials 106
4.4.3 Protein purification 106
4.4.4 AldDH fragments enzyme assays 108
4.4.5 ADH fragments enzyme assays 109
4.4.6 Fragment 11 optimum temperature assays 114
4.4.7 Fragment 11 thermo-stability assays 114
4.4.8 Estimation of the size of Fragment 11 115
4.4.9 Association experiments 117
4.4.10 Metal ion stimulation of ADH activity 117
4.5 Discussion 118

Chapter 5

Crystallisation of the minimal functional ADH domain (Fragment 11)

5.1 Introduction 121
5.2 Materials and Methods (X-ray crystallography) 122
5.2.1 Pre-crystallisation tests 123
5.2.2 Crystallisation screens 124
5.2.3 Optimisation of crystallisation conditions 124
5.2.4 Data collection 124
5.2.5 Molecular replacement and model refinement 124
5.2.6 Metal ion analysis 125
5.2.7 Thrombin cleavage of His-Tag 126
5.3 Unsuccessful crystallisation trials 126

5.4 Fragment 11 crystallisation results 127
5.4.1 Protein crystallisation 127
5.4.2 Data collection and molecular replacement 128
5.4.3 Data refinement 129
5.4.4 Interpretation of the Fragment 11 crystal structure 130
5.4.5 N-terminal domain modelling work 153
5.5 Discussion 162

Chapter 6

Identification & characterisation of other functional acetylating aldehyde dehydrogenases from *Geobacillus thermoglucosidasius*

6.1 Introduction 170
6.2 Materials and methods 171
6.3 Gene identification 171
6.3.1 AldDH protein cloning 171
6.3.2 Maltose binding protein tag 172
6.3.3 Enzyme assays 172
6.4 Results 173
6.4.1 Gene identification 173
6.4.2 Gene cloning 174
6.4.3 Protein expression trials and protein purification 175
6.4.4 Kinetic characterisation of the AcAldDH protein 177
6.4.5 Investigation into product inhibition of AcAldDH domain 180
6.4.6 Preliminary substrate specificity experiments 183
6.4.7 AcAldDH thermostability assays 184
6.4.8 AcAldDH optimum temperature assays 184
6.4.9 Oligomeric nature of AcAldDH 185
6.5 Discussion 186

Chapter 7

Crystallisation of the acetylating aldehyde dehydrogenase protein (AcAldDH)

7.1 Introduction 190
7.2 AcAldDH methods 190
7.2.1 Soaking experiments 191
7.2.2 Room-temperature diffraction 191
7.3 AcAldDH crystallisation results 191
7.4 AcAldDH protein crystallisation without substrates 192
7.4.1 Crystallisation conditions 192
7.4.2 Data collection and molecular replacement 193
7.4.3 Data refinement 193
7.4.4 Interpretation of the AcAldDH structure 196
7.4.5 Other molecules in the structure 198
7.4.6 Multimeric assembly 200
7.4.7 Modelled NAD^+ binding 203
7.5 AcAldDH protein co-crystallisation with substrate-product mix 205
7.5.1 Crystallisation conditions 205
7.5.2 Data collection and molecular replacement 206
7.5.3 Data refinement 207
7.5.4 Overview of the co-crystallised structure 208
7.5.5 Regions of additional electron density 210
7.6 Discussion 213

Chapter 8
Creation of novel artificial ADHE proteins
8.1 Introduction 218
8.2 Materials and methods 219
8.2.1 Fusion 1 cloning 220
8.2.2 Fusions 2-4 cloning 220
8.2.3 TM400 Fusion 1 vs ADHE fermentation run 221
8.3 Results 221
8.3.1 Fusion 1 221
8.3.2 Fusion 1 crystallisation trials 230
8.3.3 Fusions 2-4 230
8.3.4 Fusions 1 & 2: *in vivo* evaluation 236
8.4 Discussion 242
Chapter 9
Investigation into the expression of key metabolic enzymes through a fermentation run
9.1 Introduction 246
9.2 Materials and methods 247
9.2.1 Enzyme assays 247
9.2.2 Fermentation run (TM444) 250
9.3 Results 251
9.3.1 TM242 kinetic parameter determination (aldDH and PAT) 251
9.3.2 TM444 fermentation profiles 253
9.3.3 Specific activity measurements 254
9.3.4 Percentage activity measurements 254
9.3.5 PDH expression 255
9.3.6 AldDH and ADH expression 256
9.3.7 PAT and AK expression 257
9.4 Discussion 257
Chapter 10
General discussion and future work
10. General discussion and future work 262
Chapter 11
References
11. References 267
Appendix 1 276

ABSTRACT

Unlike first generation biofuels, those produced from ligno-cellulosic waste material (second generation) have the potential to offer sustainable fuel production without competition for food products, whilst making significant savings in terms of greenhouse gas emissions. Second generation bioethanol has the potential to offer a stop-gap between current vehicle fuelling technologies and future solutions such as biohydrogen.

TMO Renewables Ltd, a leading developer of the second-generation conversion of biomass to biofuel, has engineered the organism *Geobacillus thermoglucosidasius* to optimise its production of ethanol. This thermophilic bacterium grows optimally at 60-65°C, on a wide range of different substrates including both C_5 and C_6 sugars. The enzyme responsible for ethanol production has been shown to be a highly-expressed bifunctional enzyme (ADHE) that possesses both an acetylating aldehyde dehydrogenase (aldDH) and an alcohol dehydrogenase (ADH) activity. This enzyme is responsible for catalysing the reduction of acetyl-CoA to ethanol via an acetaldehyde intermediate:

$$\text{acetyl-CoA} + \text{NADH} + \text{H}^+ \rightarrow \text{acetaldehyde} + \text{CoA-SH} + \text{NAD}^+$$
$$\text{acetaldehyde} + \text{NADH} + \text{H}^+ \rightarrow \text{ethanol} + \text{NAD}^+$$

Here we report the characterisation of the bifunctional ADHE in terms of catalytic activity, substrate promiscuity and multimeric assembly. The properties of this enzyme in relation to competing pathways in fermentative metabolism, including its expression pattern during fermentation, have also been determined.

Investigations included the sub-cloning and separate recombinant expression of the aldDH and ADH halves of the protein, followed by the determination of a high-resolution crystal structure of the active ADH domain. The structure of the aldDH domain has been modelled, including its interaction with the ADH component of ADHE.

An additional aldDH gene was identified in *Geobacillus thermoglucosidasius;* this has also been cloned and expressed, and the recombinant enzyme characterised and its high-resolution crystal structure determined.

Through fusion of ADH and aldDH genes, a series of novel ADHE enzymes have been generated and their effect on ethanol production within the engineered *G. thermoglucosidasius* strain evaluated.

LIST OF ABBREVIATIONS

Abbreviation	Definition
#	Number
Δ	Variable
2-SPY-NG	Soy peptone yeast extract
AcAldDH	Acetylating aldehyde dehydrogenase
acs	Acetyl-CoA synthetase
ADH	Alcohol dehydrogenase
ADHE	Bifunctional aldDH-ADH protein
AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
AK	Acetate kinase
AldDH	Aldehyde dehydrogenase
ALDH	$NAD(P)^+$-dependent aldehyde dehydrogenase
ArcX	Arctic express
BSA	Bovine serum albumin
Carb	Carbenicillin
Chlor	Chloramphenicol
CoA-SH	Coenzyme A
DHQ_Fe-ADH	Dehydroquinate synthase-like (DHQ-like) and iron-containing alcohol dehydrogenases
DLS	Dynamic light scattering
dNTPs	Deoxyribonucleoside triphosphates
DTNB	5,5′-dithiobis-(2-nitrobenzoic acid)
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EC	Enzyme commission (number)
EDTA	Ethylenediaminetetraacetic acid
EPPS	4-(2-Hydroxyethyl)-1-piperazinepropanesulfonic acid
EutE	Ethanolamine utilisation protein E
F	Structural factor
F_c	$F_{calculated}$
F_o	$F_{observed}$
FPLC	Fast protein liquid chromatography
Gent	Gentamycin
GHG	Greenhouse gas
GSH	Reduced glutathione
h	Hours
His-tagged	Histidine tagged
HPLC	High performance liquid chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Kelvin
k/o	Knock-out
Kan	Kanamycin
kb	Kilo-bases
K_i	Dissociation constant of EI
K_i'	Dissociation constant of ESI
K_m	The Michaelis constant (concentration of substrate when reaction rate is ½ of V_{max})
LB	Lysogeny broth
ldh	Lactate dehydrogenase
MALDI-TOF/TOF	Matrix assisted laser desorption/ionisation-time of flight/time of flight

MilliQ H_2O	Ultrapure water
min	Minutes
M_r	Relative molecular mass
nm	Nano meters
NTB^{2-}	2-nitro-5-thiobenzoate^{2-}
OD	Optical density
pat	Phosphate acetyl transferase
PCR	Polymerase chain reaction
pdh	Pyruvate dehydrogenase
pfl	Pyruvate formate lyase
PGA-LM	Poly-glutamic acid
pI	Isoelectric point
PTA	Phosphotungstic acid
s	Seconds
SAP	Shrimp alkaline phosphatase
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEM-EDS	Scanning electron microscope-energy dispersive X-ray spectroscopy
SOC	Super optimal broth with catabolite repression
Strep	Streptomycin
TAE	Tris-acetate-EDTA
Taq	*Thermus aquaticus*
TB	Terrific broth
TEM	Transmission electron microscopy
TMO	Thermophilic micro-organism
T_{opt}	Optimum temperature
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
U	µmol/min
USM	Urea sulphate media
V_{max}	Maximum velocity of reaction
X-Gal	5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside
YENB	Yeast extract nutrient broth
ε	Extinction coefficient

1 INTRODUCTION

1.1 Energy demands

The world economy is dependent on the production and transfer of energy and on human labour. According to the international energy agency in 2009, over 8000 Mtoe (million tonne oil equivalent) were used worldwide (IEA 2011). The general trend observed is an increase in energy usage since accurate records began. In 2010 more than 80% of the energy consumed by the global economy was derived from finite sources such as oil, natural gas, peat and coal (The World Bank Group 2012). As shown in Figure 1-1, most countries are heavily dependent on the use of energy derived from fossil fuels. The combustion of such fuels is the widely accepted cause of increasing carbon dioxide and other greenhouse gas (GHG) emissions, and a major causative factor in global climate change. The continuing development of previously less-developed countries is likely to result in increases in worldwide energy usage in the future.

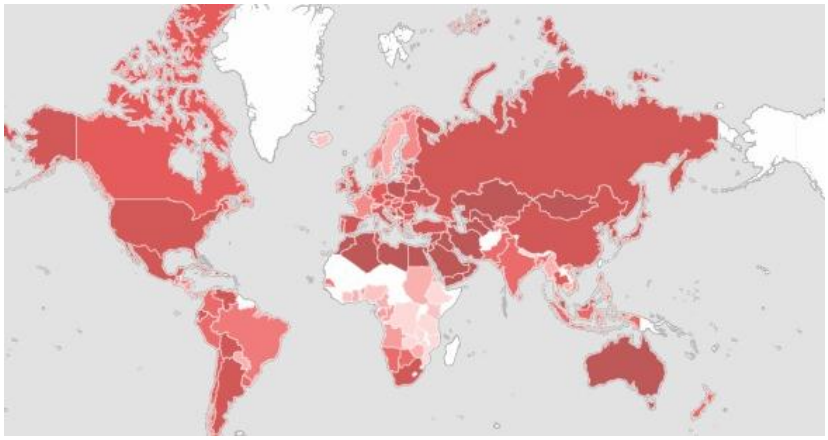


Figure 1-1: World map coloured by % fossil fuel energy of total energy consumption from 2007-2011 (coloured scale: white = 54.1% and darkest red shown = 68.1% fossil fuel energy use) (The World Bank Group 2012).

A major contributor to global GHG emissions is transportation. The Department for Transport report that over 34 million motor vehicles were registered in 2011 for use in the UK alone (Department for Transport 2012). In 2009, 27% of GHG emissions produced in the UK originated from international and domestic transport (Department for Transport 2011), and this equates to 122 million tonnes of CO_2 equivalent ($MtoCO_2e$). Although improvements in engine efficiency and a reduction in domestic travel due to current economic factors have seen this figure decrease in recent years, consumption of fossil fuels for transportation remains a significant environmental issue.

The wide variety of uses of crude oil derivatives other than for fuel means that consumption of this finite resource by combustion is becoming undesirable. Therefore, the search for renewable alternatives to current transportation fuels has become increasingly important in recent years.

Biohydrogen (see Section 1.2) and renewable electricity derived from solar, hydroelectric and wind energy, are potential sources of renewable energy that may serve as long-term, clean alternatives to oil-based transport technologies. The current dependence on liquid fuel technology (i.e. petrol and diesel) means that many alternative fuel sources for transportation would require extensive redevelopment of current vehicle fuelling infrastructure, which may prove prohibitively expensive in the short term. Limitations in vehicle range and sustainability with these fuel technologies mean that a short-term renewable solution is required. Biofuels such as bioethanol and biodiesel may provide a "stop-gap" solution between current liquid-based fuelling technology and future fuel systems.

1.2 Biofuels

Fuel products that can be derived from various different forms of biomass are defined as biofuels. These fuels present a renewable energy source derived from the sun's energy, most commonly through photosynthesis by plants. Combustion of fossil fuels results in the release of previously sequestered CO_2 that has not been present in the atmosphere for millions of years. In the case of biofuels, atmospheric CO_2 sequestered during the growth of the plant material, is released when the derived biofuel is combusted. These fuels therefore have the potential to offer a significant saving in GHG emissions when compared to fossil fuels.

Major biofuel products include bioethanol, biobutanol, biodiesel, biohydrogen and methane. A wide range of possible biomass feedstocks can be used to produce this range of biofuel products (Barnard et al. 2010; Luque et al. 2008). Some of these processes are currently driven by chemical conversion, for example biodiesel, whereas others are carried out using microorganisms to convert biomass material, for example methane and bioethanol (Antoni et al. 2007).

Two major biofuels of current industrial significance are bioethanol and biodiesel; these represent more than 90% of the world biofuel market (Antoni et al. 2007). Biodiesel is produced from various oils and fats (such as sunflower or soybean oils) by a process of transesterification. Triglycerides are reacted with an alcohol, most commonly methanol

(with an alkaline catalyst), to produce monoalkyl esters, which can be used in diesel-powered engines with few modifications (Barnard et al. 2010; Meher et al. 2006). Glycerol is a by-product of this reaction. The majority of bioethanol production is focussed around traditional fermentation techniques. These use microorganisms, such as the yeast *Saccharomyces cerevisiae,* to convert simple sugars from pre-treated feedstocks, such as corn or sugar cane, to yield ethanol (Sanchez and Cardona 2008). Conventional petrol-fuelled vehicles can run on 5% ethanol inclusion fuel with no modifications. Vehicles can be converted to use up to 85% ethanol using relatively inexpensive products that are already commercially available (Department for Transport 2010; Fuel Flex International TM 2012). Conversion of vehicles to usage of high biofuel content fuels could represent a significant reduction in oil consumption, and if sourced responsibly a reduction in GHG emissions.

Although first generation biofuels produced from feedstocks such as corn, wheat, soybean oil and sugar cane are well established technologies, they have been well publicised as a potential cause of global food price increases in the "food vs fuel debate". In the year 2008-2009, 37% of the US corn harvest was used for ethanol production (Martin 2010). The FAO predict that a 34% increase in world population size by 2050, coupled with increasing urbanisation of previously less-developed areas, may result in a 70% increase in global food demand (Food and Agriculture Organization of the United Nations 2009). The use of first-generation biofuels may contribute further by reducing the supply of crops for food. They have also been highlighted as potentially having severe disadvantages where land is converted for energy crop production, where a net GHG emission saving may not be made for many years (Fargione et al. 2008; Searchinger et al. 2008). Energetic requirements and environmental impacts of a particular feedstock must be accurately determined, to ensure the production of the biofuel product does indeed have a net energy gain over its production energy costs. Hill et al (2006) suggested that the overall saving in GHG emissions for corn-based bioethanol and soybean-based biodiesel were 12 and 43% respectively. The net energy gain for these two products was 25% for corn-based bioethanol and 93% for soybean-biodiesel (Hill et al. 2006). More recently, others have suggested that once land usage changes are also taken into account, some biofuels can actually represent a net increase in GHG emissions (Searchinger et al. 2008).

In contrast to first-generation biofuels, second-generation biofuel technologies focus on the use of non-food feedstocks to generate fuel products. Examples include bioethanol produced from municipal waste, ligno-cellulosic waste materials (such as corn husks,

cassava root and wheat straw) or bioenergy crops (such as switchgrass or poplar trees), and the use of microalgal species with a high level of intracellular oils, for triglyceride production for biodiesel (Luque et al. 2008). The use of second-generation bioethanol in the transesterification reaction would further improve the sustainability of the biodiesel technology. These technologies have also been suggested to provide significantly higher savings in GHG emissions (Searchinger et al. 2008). This is illustrated in Figure 1-2, particularly when produced from waste material where land usage changes are not apparent.

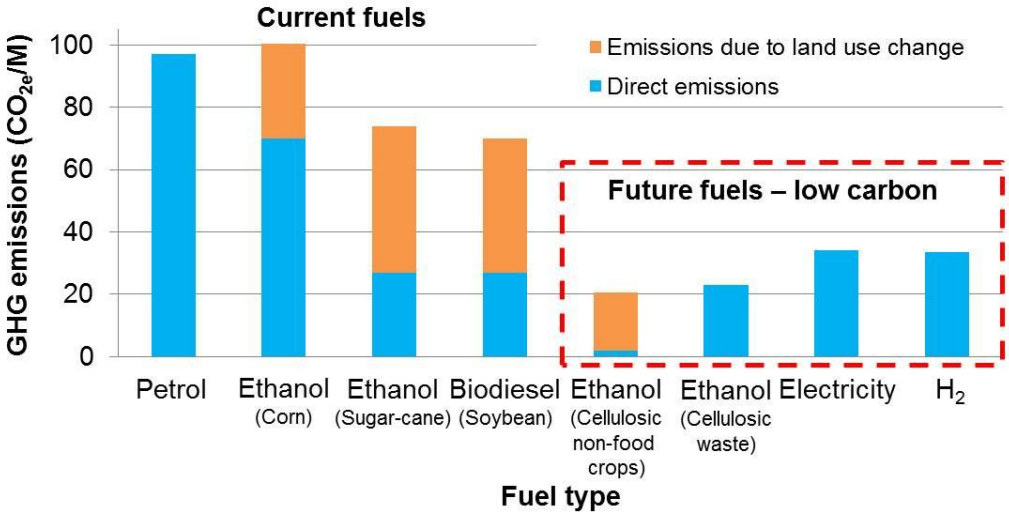


Figure 1-2: Summary of GHG emissions for current and predicted fuels determined for the California Air Resources Board (Charles 2009). Image based on figure from *Science* 1 May 2009: Vol. 324 p. 587.

It should be emphasised that it is unlikely that one technology can provide the solution to the world's low carbon energy requirement. Due to limitations in feedstock supplies and environmental factors, a range of solutions, including various biofuel technologies, will be required to maintain and extend access to energy that is likely to be demanded in the future (Potocnik 2007).

This project is in part sponsored by a second-generation bioethanol company, TMO Renewables Ltd, and so this specific technology will become the focus of the rest of this Introduction.

1.3 Second-generation bioethanol

As discussed previously, the use of non-food biomass for the production of second-generation bioethanol has many environmental advantages. The feedstocks associated with this technology are derived from plant material such as wheat straw, cassava roots and wood chips. The resistant nature of these feedstocks to

degradation means that development of this technology is heavily dependent on pre-treatment steps. This treatment is required to liberate sugars from the complex lignin-cellulose-hemicellulose matrix that makes up plant cell walls (Lynd et al. 2008).

The composition of plant cell walls varies between feedstocks but is broadly made up of cellulose strands (a polymer of D-glucose molecules) cross-linked by hemi-cellulose (includes D-xylose, D-mannose, D-galactose and L-arabinose polymers); the matrix is further strengthened by lignin and pectin. This complex matrix of interlinked polymers requires both physical and biochemical treatment to release fermentable sugars for bioethanol production (Jordan et al. 2012; Sanchez and Cardona 2008). Efficient pre-treatment of biomass typically employs a physical step such as steam explosion (with acid hydrolysis), followed by an extensive cellulolytic enzyme treatment to degrade the partially-disrupted polymeric sugars. The costs associated with the addition of these enzymes and effective physical pre-treatment can have a significant impact on the overall cost of the bioethanol produced (Lynd et al. 2008).

The sugars released by this pre-treatment are diverse and can include short-chain polysaccharides (cellobiose and xylobiose) as well as monosaccharaides, including both C_6 sugars (D-glucose and D-mannose) and C_5 sugars (D-xylose and L-arabinose). An array of other compounds can also be produced from minor cell wall components. Traditional fermentation techniques using organisms such as *S. cerevisiae* are only able to utilise the simple C_6 sugars to produce ethanol. However, C_5 sugars can form a significant proportion of the sugars present in the pre-treated material (Limayem and Ricke 2012). Increasing the proportion of biomass material that can be utilised is of fundamental importance if second-generation bioethanol is to become an economically-viable alternative to petrol. Extensive research is therefore being undertaken by various groups in an attempt to extend the substrate promiscuity of these organisms (Barnard et al. 2010; Jordan et al. 2012; Sanchez and Cardona 2008).

A possible solution to the narrow substrate range of traditional fermentative organisms is to use novel ethanologenic species with broader substrate capacity. This approach also has potential limitations in terms of organism ethanol tolerance and yields, limited metabolic manipulation tools and the potential requirement for redevelopment of fermentation methodologies.

Thermophilic microorganisms such as some species of *Clostridia, Thermoanaerobacter* and *Geobacillus* have been identified as suitable for use as industrially-viable ethanol

producing organisms (Barnard et al. 2010; Sanchez and Cardona 2008; Taylor et al. 2008). Potential advantages of using thermophilic organisms include:

- broad fermentable substrate specificity (C_5 and C_6 sugars)
- robust metabolic systems less sensitive to fluctuations in temperature and pH
- lower contamination risks (due to higher growth temperatures)
- high temperature fermentation may allow process optimisation (i.e. less energy in cooling hydrolysate and facilitation of ethanol removal).

1.4 *Geobacillus thermoglucosidasius*

The second-generation biofuel company TMO Renewables (Guildford UK) selected the thermophilic microorganism, *Geobacillus thermoglucosidasius* NCIMB 11955, as a suitable candidate organism for sustainable bioethanol production. This Gram-positive bacterium, which grows optimally between 60-65°C, was selected due to the diverse range of substrates on which it could grow, as well as its ability to produce low levels of ethanol. The strain is able to utilise C_5 and C_6 sugars, including some oligomeric forms derived from biomass feedstocks, to produce lactate, formate, acetate and ethanol.

Over recent years, the company have engineered the wild-type strain to optimise its metabolism in terms of ethanol production. This required a series of gene knock-outs and up-regulations to significantly increase the yield of ethanol obtained when grown anaerobically; these are outlined in Figure 1-3. This process utilised a “genetic toolbox” developed by TMO Renewables to allow manipulation of various genes and the controlling promoters (Cripps et al. 2009).

When *G. thermoglucosidasius* is grown anaerobically, its metabolism concerned with ethanol production is driven by the requirement to regenerate NAD^+ from NADH. This reduced form of the cofactor is produced by a variety of metabolic processes such as glycolysis. Pyruvate, a product of the pentose-phosphate, Entner-Doudoroff and the Embden-Meyerhof pathways, is the major substrate for fermentative metabolism. In the wild-type organism the dominant process to regenerate NAD^+ is the action of the lactate dehydrogenase enzyme, which does not lead to ethanol production. TMO Renewables removed the gene responsible for this activity to force the dominant activity to shift towards ethanol production. The metabolic flux through two different enzymes to form acetyl-CoA therefore increased. These enzymes are the pyruvate formate lyase and the pyruvate dehydrogenase complex (PDH). The dominant activity in the wild-type organism was that of the pyruvate formate lyase. However, the

accumulation of the organic acid formate that results from this enzyme activity inhibited the growth of the organism; therefore this enzyme was subsequently knocked out, directing the flux of carbon through PDH. As the expression of this enzyme was low under anaerobic conditions in the wild-type organism, this required up-regulation to prevent this enzyme becoming rate limiting. This was achieved by placing the gene under the control of a promoter active under oxygen-limited conditions (the lactate dehydrogenase promoter from *G. stearothermophilus*).

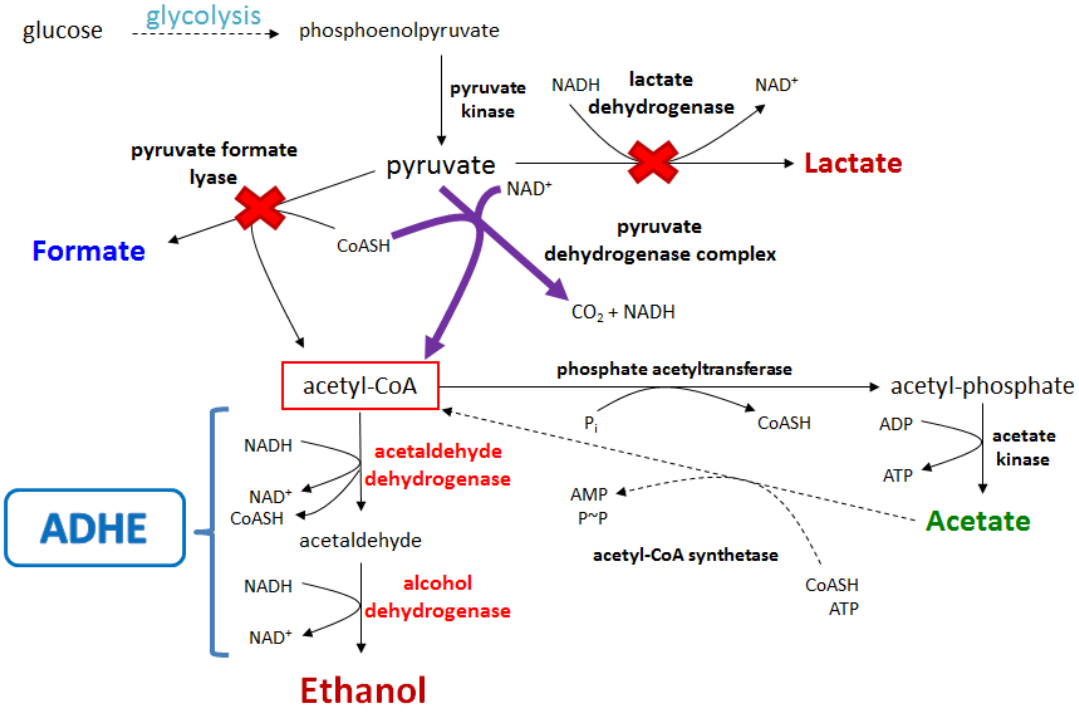


Figure 1-3: Overview of modified fermentative metabolism of the TM242 (and TM444) strain. Red crosses indicate gene knock-outs and purple arrows indicate up-regulations through promoter switching. A hypothesised acetyl-CoA synthetase activity is shown with a dashed arrow. The two dehydrogenase activities associated with the ADHE protein are indicated.

Acetyl-CoA, the product of PDH, is a substrate for two relevant enzymes: phosphate acetyl transferase (PAT) and an aldehyde dehydrogenase (aldDH) enzyme. The product of PAT (acetyl-phosphate) is converted to acetate by the action of acetate kinase (AK). This branch of metabolism has not currently been successfully knocked out, with PAT and AK gene knock-outs proving lethal; this suggests an essential function in the organism. This essential function may be to regenerate ATP for the organism to grow anaerobically. The product of the aldDH enzyme (acetaldehyde) is converted to ethanol by the action of an alcohol dehydrogenase enzyme (ADH). Of a possible 8 annotated ADH genes within the TM242 genome, only one gene when knocked out completely abolished ethanol production. This was a putative *adhE* gene

hypothesised to possess both ADH and aldDH activity. It is around this protein that this project is focussed.

Two molecules of NADH are generated per acetyl-CoA, one by glyceraldehyde 3-phosphate dehydrogenase during glycolysis, and the second by PDH. The ethanol yielding activity of the two dehydrogenases associated with the ADHE protein is therefore required during fermentative metabolism to maintain the redox balance of the cell. The metabolic manipulations described produced the modified strain known as TM242. This was the production strain used by TMO Renewables at the outset of this project. Greater than 90% of the theoretical yield of ethanol was obtained when this modified strain was fermented on glucose. The strain also preformed promisingly on other sugars that can be derived from biomass including cellobiose, D-xylose and L-arabinose (Cripps et al. 2009). The strain has now been further modified to remove the ability of the organism to sporulate, thereby producing the current production strain TM444.

Strain TM444 forms the basis of the TMO Renewables process that is currently being commercialised. The company have constructed an industrial-scale development plant in the UK to demonstrate to potential clients the commercial viability of bioethanol production from various feedstocks. The first commercial agreement was signed between TMO Renewables and the US waste management firm Fiberight in 2010. Commercial-scale plants will use the process to produce bioethanol from household waste (Barley 2010).

1.5 ADHE

The single protein hypothesised to catalyse the conversion of acetyl-CoA to ethanol in *G. thermoglucosidasius* is known as ADHE. Preliminary evaluation of an ADHE knock-out strain of TM242 (TM400) showed complete abolition of ethanol production and an inability of the strain to survive under anaerobic conditions (TMO Renewables 2009 personal communication). Knock-outs and down regulations of the *adhE* gene have been shown to negatively affect the phenotype of solvent production in several organisms including *Escherichia. coli* (Goodlove et al. 1989), *Entamoeba histolytica* (Espinosa 2001), *Giardia lamblia* (Dan and Wang 2000) and *Clostridium acetobutylicum* (Fontaine et al. 2002).

AdhE genes are common amongst fermentative microorganisms and have been studied in a range of different organisms, as shown in Table 1-1. The protein is

commonly associated with fermentative metabolic pathways catalysing the conversion of an acyl-CoA to an alcohol. The most common activity described is conversion of acetyl-CoA to ethanol, although in some *Clostridial* sp it has also been shown to be capable of catalysing conversion of butyryl-CoA to butanol (Fontaine et al. 2002).

A significant amount of research has focussed on the ADHE proteins from *E. coli* and *En. histolytica.*

Type of organism	**Species**	**Reference**
Gram-positive bacteria	*Leuconostoc mesenteroides*	(Koo et al. 2005)
	Thermoanaerobacter ethanolicus	(Peng et al. 2008)
	Thermoanaerobacterium saccharolyticum	(Shaw et al. 2008)
	Lactococcus lactis	(Arnau et al. 1998)
	Streptococcus bovis	(Asanuma et al. 2004)
	Clostridium acetobutylicum	(Fontaine et al. 2002)
Gram-negative bacteria	*E. coli*	(Membrillo-Hernandez et al. 2000)
Anaerobic protozoa	*Giardia lamblia*	(Dan and Wang 2000)
	Entamoeba histolytica	(Bruchhaus and Tannich 1994)
Algae	*Polytomella* sp (within the mitochondrion)	(Atteia et al. 2003)

Table 1-1: A selection of organisms where ADHE proteins have been studied.

The *adhE* gene sequence annotated in the TM242 genome was translated during this project to provide the amino acid coding sequence using the Translate tool (ExPASy (Gasteiger E. 2005)) (Section 2.3). A protein BLAST search (NCBI) indicates that the protein is made up of two distinct protein folds linked together. The N-terminus of the protein aligns with an aldehyde dehydrogenase family of proteins, whereas the C-terminal domain aligns with an iron containing ADH family (Figure 1-4). This is the characteristic arrangement seen for other ADHE proteins with the two dehydrogenase domains linked in a single “fusion” protein.

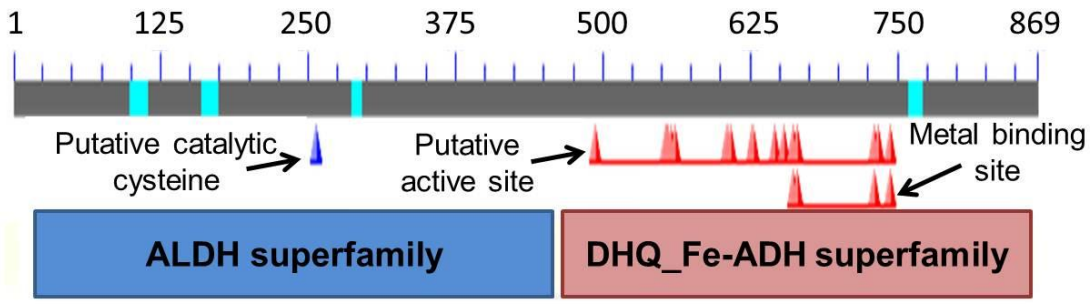


Figure 1-4: Result summary from a protein BLAST of the TM242 ADHE sequence against the NCBI database. Numbers correspond to amino acid residues. Light blue regions indicate compositionally biased regions that were excluded from the search by the programme. Putative catalytic sites are indicated. ALDH = $NAD(P)^+$-dependent aldehyde dehydrogenase superfamily and DHQ_Fe-ADH = Dehydroquinate synthase-like (DHQ-like) and iron-containing alcohol dehydrogenases superfamily.

Catalysis of the conversion of acetyl-CoA to ethanol can in theory be carried out by this single protein unit, where the aldehyde product of the first enzyme is consumed by the second. It is not clear whether there is an evolutionary benefit to the expression of these two activities on a single polypeptide. It can be envisaged that accumulation of the intermediate product (acetaldehyde) may have a negative effect on the fermenting organism, so efficient conversion to ethanol may be desirable. The ability of acetaldehyde to react with proteins within the cell means it is considered a toxic compound. Substrate channelling between the two domains may allow rapid conversion of the intermediate product, limiting the accumulation within the cell. Fusion of these two domains may allow positioning of the active sites to facilitate this. The two dehydrogenase activities are required to maintain the redox balance in fermentative organisms; it may therefore prove advantageous to express these two activities together to optimise cofactor regeneration.

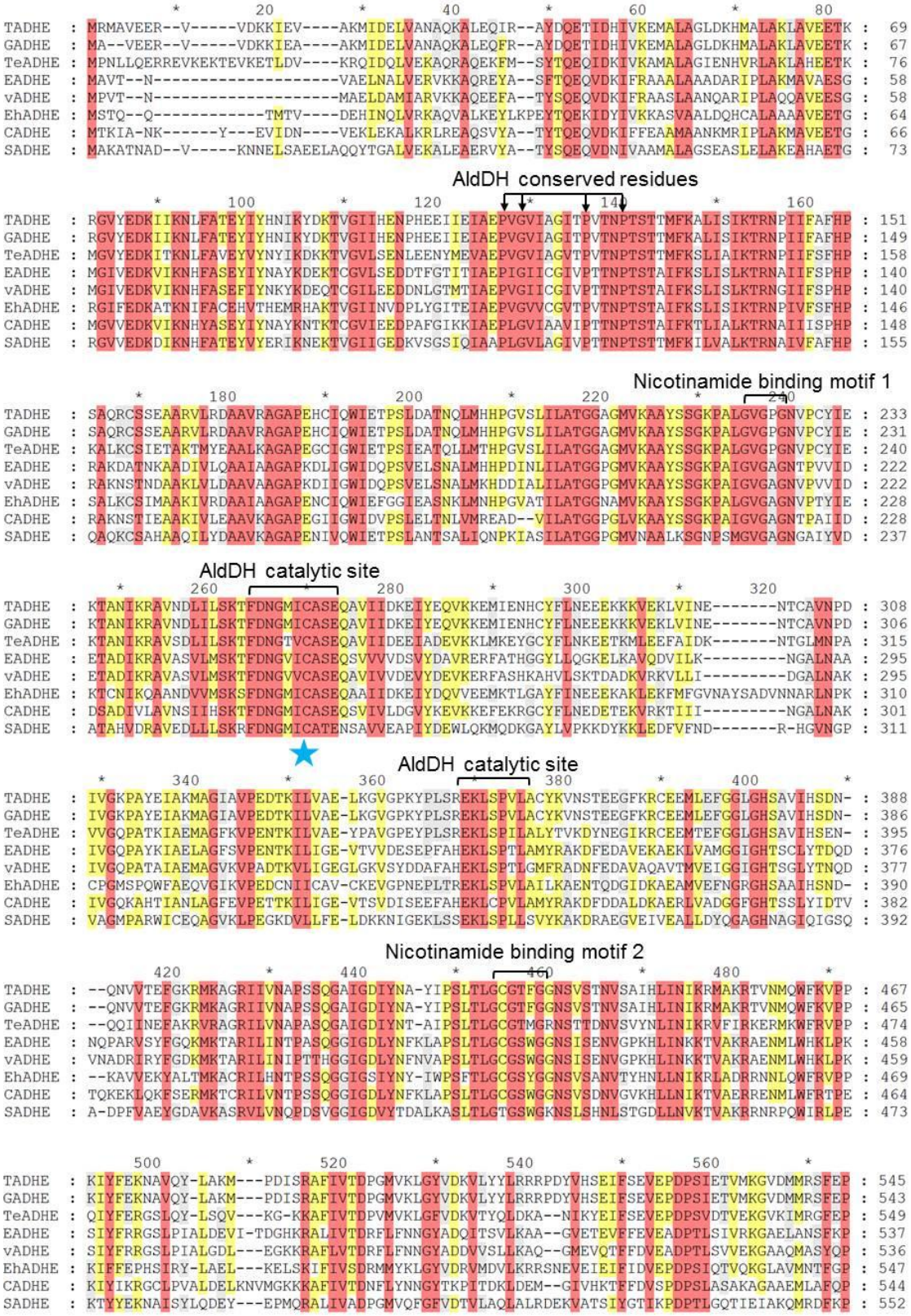
1.5.1 Sequence analysis

Sequence alignments between the ADHE from TM242 and other ADHE sequences carried out here show that the protein is well conserved between organisms, with over 45% sequence identity being observed for all sequences (Table 1-2). The ADHE protein sequence from TM242 appears to be 99% identical to that from the closely related *G. thermoglucosidasius* C56-YS93 strain. To the author's knowledge, no characterisation of ADHE proteins from *G. thermoglucosidasius* or closely related strains has been carried out.

Protein	G_{ADHE}	Te_{ADHE}	E_{ADHE}	V_{ADHE}	Eh_{ADHE}	C_{ADHE}	S_{ADHE}
ID %	99	66	47	47	57	49	47
Similarity %	99	82	67	67	74	68	65
Gaps %	0	1	5	6	2	2	3

Table 1-2: Protein sequence alignment scores between representative ADHE proteins and ADHE from TM242. G_{ADHE} = *Geobacillus thermoglucosidasius* C56-YS93 ADHE (AEH49709.1), Te_{ADHE} = *Thermoanaerobacter ethanolicus* ADHE (ABH06551.1), E_{ADHE} = *E.coli* K-12 ADHE (NP_415757.1), V_{ADHE} = *Vibrio parahaemolyticus* RIMD 2210633 ADHE (NP_798500.1), Eh_{ADHE} = *Entamoeba histolytica* ADHE (Q24803), C_{ADHE} = *Clostridium thermocellum* ATCC 27405 ADHE (YP_001036854.1), S_{ADHE} = *Streptococcus bovis* ADHE (BAC87790.1). Numbers in brackets indicate protein accession numbers. Alignments were carried out using the T-Coffee sequence alignment tool (ExPASy) (Notredame et al. 2000) and interpreted using the GeneDoc software package (Nicholas et al. 1997).

Several key residues and motifs are conserved between the proteins in the alignment. Figure 1-5 shows the annotated alignment using the PROSITE tool to predict nicotinamide- and metal-binding regions (ExPASy) (Sigrist et al. 2010); conserved motifs discussed by Fontaine et al (2002) are also shown. The aldDH domain contains the highly conserved aldDH residues and the catalytic site residues predicted for similar proteins. The ADH domain contains a metal ion binding motif. Four hypothetical nicotinamide binding regions were identified within the coding sequence.


AldDH conserved residues
Nicotinamide binding motif 1
AldDH catalytic site
AldDH catalytic site
Nicotinamide binding motif 2

Continued…

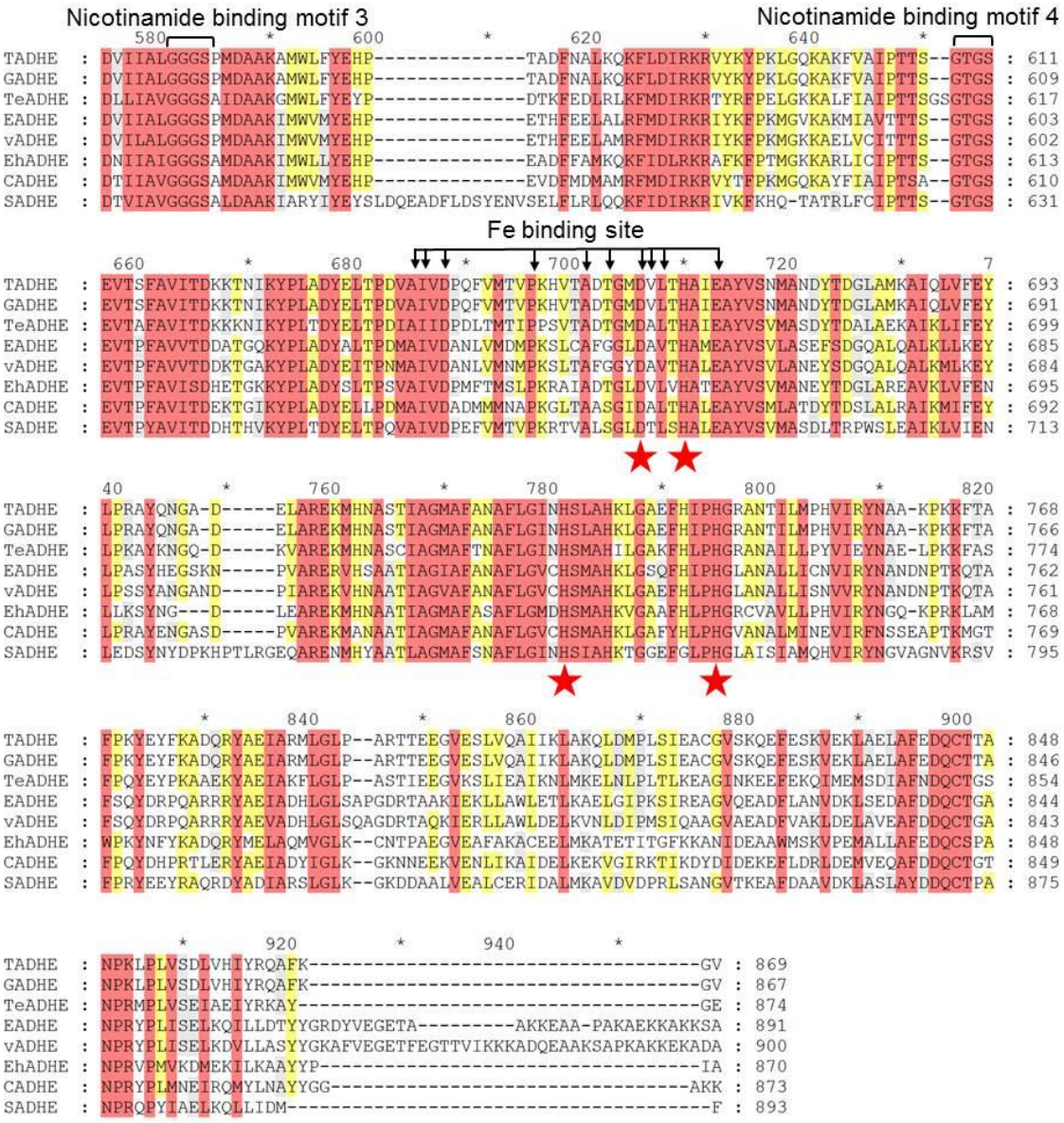


Figure 1-5: Protein sequence alignment between representative ADHE proteins and ADHE from TM242. Residues highlighted in red are well conserved between all aligned sequences. Yellow highlighted are conserved or conservative changes in most of the sequences, grey residues are only partially conserved and conservative changes, whereas white indicates less conservation. TADHE = TM242 ADHE, GADHE = *Geobacillus thermoglucosidasius* C56-YS93 ADHE (AEH49709.1), TeADHE = *Thermoanaerobacter ethanolicus* ADHE (ABH06551.1), EADHE = *E.coli* K-12 ADHE (NP_415757.1), VADHE = *Vibrio parahaemolyticus* RIMD 2210633 ADHE (NP_798500.1), EhADHE = *Entamoeba histolytica* ADHE (Q24803), CADHE = *Clostridium thermocellum* ATCC 27405 ADHE (YP_001036854.1), SADHE = *Streptococcus bovis* ADHE (BAC87790.1). Nicotinamide and iron binding motifs were identified using the PROSITE tool (ExPASy). Blue star below the sequence indicates conserved catalytic cysteine identified during BLAST search. Red stars below the sequence indicate conserved metal ion coordinating residues identified from ADH domain crystal structure determined during this project. AldDH conserved residues were based on those sequences discussed by Fontaine et al (2002). Numbers in brackets indicate protein sequence accession numbers. Alignment carried out using the T-Coffee sequence alignment tool (ExPASy) (Notredame et al. 2000) and interpreted using the GeneDoc software package (Nicholas et al. 1997).

Conserved residues within the aldDH domain of ADHE have previously been identified in other protein sequence alignments (Chen et al. 2004; Fontaine et al. 2002; Perozich et al. 1999). A catalytic cysteine residue is highly conserved within aldDH proteins and is also present in the TM242 ADHE protein sequence (Cys_{257}); this is indicated in Figure 1-5 with a blue star. A glycine (Gly_{254}) at position -3 from this residue is highly conserved as well as a glutamic acid (Glu_{348}) in an EKLSP motif that appears in the aligned protein sequences. Mutations in the catalytic cysteine and the equivalent glutamic acid in the *E. histolytica* ADHE protein have been shown to abolish aldDH activity (Chen et al. 2004). The conserved proline residues described by Fontaine et al (2002) are also conserved in the sequence alignment performed here. Analysis of the amino acid sequence using the PROSITE tool revealed one conserved $NAD(P)^+$ binding motif within the aldDH domain, two $NAD(P)^+$ binding motifs within the ADH domain, as well as one motif positioned between the two domains of the ADHE protein.

Analysis of the ADH domain revealed an iron-containing ADH motif 1 (Fe binding site) within the protein sequence that is conserved between the aligned ADHE proteins. This ADH family is distinguished from the more commonly described zinc dependent ADH superfamily in terms of sequence homology. Three histidine residues (His_{665}, His_{730} and His_{744}) and an aspartic acid (Asp_{661}) shown to be coordinating the binding of a metal ion during crystallographic study of the ADH domain during this project (Chapter 5), were also shown to be highly conserved. The metal ion is hypothesised to be catalytic, polarising the acetaldehyde carbonyl group allowing the formation of the alcohol group (Atteia et al. 2003). In *E. histolytica*, when His_{730} or His_{744} were mutated to an arginine, both ADH and aldDH activities were lost.

The activities of the two domains of the *E. histolytica* ADHE protein have been previously isolated independently of one another through the work of Stanley's research group (Chen et al. 2004; Espinosa 2001). This work showed low levels of aldDH activity could be resolved on a protein fragment containing amino acids 1-446; the ADH activity was resolved on a fragment containing amino acids 417-870 (Chen et al. 2004; Espinosa 2001). Interestingly, the aldDH activity was not detected on a protein fragment containing amino acids 1-532, possibly due to interference by the additional C-terminal amino acids.

The translated *G. thermoglucosidasius* ADHE amino acid sequence was analysed using the ProtParam tool (Gasteiger E. 2005) to provide estimations of the basic properties of the enzyme. The protein contains 869 amino acids with a molecular

weight of 96,290. The theoretical pI of the protein is 7.11 and the molar absorption coefficient was calculated to be 67,785 $M^{-1}cm^{-1}$ at 280 nm.

1.5.2 Substrate specificity

A broad range of substrate specificities have been reported for both the aldDH and ADH domains of ADHE proteins. The aldDH domain of the *G. lamblia* ADHE shows significant activity with acetyl-CoA (100%), propionyl-CoA (52%), isobutyryl-CoA (26%), butyryl-CoA (54%), benzoyl-CoA (6%) and succinyl-CoA (3%); bracketed values indicate relative percentage activities (Sanchez 1998). It has also been reported that the ADH domain of *E. coli* ADHE is catalytically active with un-branched primary alcohols up to 6 carbons in length; the enzyme from *Salmonella typhimurium* is capable of catalysing the conversion of up to 8 carbon un-branched primary alcohols (Dailly et al. 2000). The probable ADHE enzyme from *L. mesenteroides* also showed significant activity with acetaldehyde, propionaldehyde, butyraldehyde and isobutyraldehyde when assayed in the alcohol producing direction (Kazahaya et al. 1972).

One of the goals of TMO Renewables is to use the TM242 “platform” to produce a range of different products including other fuel alcohols such as propanol and butanol. It is therefore interesting to observe the substrate specificities of various enzymes to determine if metabolic pathways may be manipulated to produce other products. If the ADHE enzyme in this organism were to show a broad specificity in terms of acyl-CoAs and aldehydes, manipulation of metabolism to achieve this goal may be significantly simpler.

1.5.3 Relative activities of the two domains

The diverse range of assay, expression and purification conditions used during previous studies of ADHE proteins makes it difficult to predict the relative activities of the two domains. Kessler et al (1991) report similar levels of activity for the two domains for the purified enzyme of *E. coli* in the forward direction (aldDH 9.6 U/mg and 9.9 U/mg for the ADH domain), whereas Espinosa et al (2009) report significantly lower aldDH activity (7.7 mmol/min/mg) compared to ADH activity (27.9 mmol/min/mg) for the partially-purified *E. histolytica* enzyme. Pei et al (2010) report that the purified *T. ethanolicus* enzyme showed no ADH activity under physiological conditions compared to 113 U/mg for the aldDH domain. It is therefore important to determine the kinetic parameters for the TM242 enzyme to determine the relative activities of the domains under defined conditions.

The aldDH domain of ADHE has also been suggested in some organisms to be relatively unstable compared to the ADH domain (Pei et al. 2010). Oxidation of the catalytic cysteine in the active site of this domain is a possible cause of this instability.

In *Thermoanaerobacter* sp, the conversion of acetyl-CoA to ethanol through fermentative pathways has been suggested to be controlled through a balance of different enzyme expression levels, enzyme activities, and ethanol concentrations (Burdette and Zeikus 1994; Pei et al. 2010). The balance of activities between various proteins including ADHA, ADHB and ADHE, have been suggested to control the metabolic flux to ethanol within these strains. The activities of these proteins appear to regulate the production of ethanol through catalysis of the forward and reverse reactions between acetyl-CoA, acetaldehyde and ethanol. Transcription of the *adha* and *adhe* genes has been shown to vary according to ethanol concentration. Although other ADH and aldDH genes have been identified in the TM242 genome, corresponding genes to *adha* and *adhb* have not yet been identified or characterised.

1.5.4 Divalent metal ions

The activity of the *E. coli* ADHE protein has been shown to be stimulated by the presence of Fe^{2+} (Kessler et al. 1991) but not by other metal ions. This was also the case for the enzyme from *E. histolytica* and *S. bovis* (Asanuma et al. 2004; Espinosa et al. 2009); moreover, in these cases the dehydrogenase activities of ADHE were inhibited by the presence of other divalent metal ions. Binding of the Fe^{2+} ion was shown to be relatively weak through stripping experiments using treatment with the metal chelator 1,10-phenanthroline (Kessler et al. 1992). Several investigations into ADHE enzymes have not described the Fe^{2+} dependent activity, with significant activity measurements being made in the absence of metal ions (Fontaine et al. 2002; Koo et al. 2005; Pei et al. 2010; Sanchez 1998). The effect of divalent metal ions is therefore investigated as part of the characterisation of the TM242 ADHE protein performed here.

1.5.5 Regulation of expression

Expression of the well-studied ADHE protein from *E. coli* has been shown to be induced under anaerobic conditions (Chen and Lin 1991), with a 10-fold increase being reported (Membrillo-Hernandez and Lin 1999). A similar induction of expression of ADHE has been demonstrated in *L. lactis* when grown on glucose as a carbon source. Interestingly, when grown on galactose, significant aerobic and anaerobic expression was observed (Arnau et al. 1998). Various regulatory factors have been

hypothesized including catabolite and nitrate repression. Particular relevance of the NAD^+/NADH balance has been shown by manipulating the intracellular levels of these cofactors (Leonardo et al. 1996).

Dailly et al (2000) have shown that the level of ADHE activity can vary according to the carbon source used during growth of both *E. coli* and *S. typhimurium*. Expression of ADHE was solely anaerobic in both cases, with the highest ADHE activity being observed when grown on reduced substrates such as sorbitol and mannitol. Low activity was observed during growth on the oxidised sugar acids such as glucuronic acid. The requirement to balance redox potential through fermentative metabolism is likely to drive the production of the ADHE protein. Metabolism of the oxidised sugar acids results in no net increase in NADH, hence the low levels of ADHE activity. Metabolism of the reduced substrates (i.e. mannitol) can result in 3 NADHs accumulating, thus requiring a higher ADHE activity (Dailly et al. 2000; Wolfe 2005). A link between *adhE* gene expression and the intracellular redox balance of NADH and NAD^+ is a possible means of regulation of the ADHE activity dependent on metabolic substrates.

Expression of the *adhE* gene in *G. thermoglucosidasius* has been suggested to be under the control of an anaerobic regulation system known as a Redox repressor (REX) (Brekasis and Paget 2003; Sickmier et al. 2005; Wang et al. 2008). Repression of the target genes by REX is achieved through the binding of the protein to the appropriate DNA operator sequence and preventing expression. It is hypothesised that the balance between NADH/NAD^+ is used to regulate the ability of this transcription factor to bind to DNA. When the cellular NADH levels are high (anoxic), the expression of the target genes is free from repression; conversely, when NAD^+ levels are higher the genes are repressed. NADH binding to the REX protein causes conformational changes that prevent its binding to DNA, and thus allows expression of the target genes (Wang et al 2008). A REX coding sequence has been identified in the TM242 genome suggesting this regulatory protein is present in this organism. A REX operator sequence is also annotated in the ADHE promoter region (TMO Renewables 2012 personal communication). Preliminary investigations have confirmed the binding of REX to the promoter region of ADHE in *G. thermoglucosidasius* (Paget, M. S. 2012 unpublished work).

An ADHE-regulating, redox-sensing protein known as RSP has been identified in *Thermoanaerobacter* spp (Pei et al. 2011). This protein showed 43% identity to the

REX protein from *Thermus aquaticus* and appears to play a similar role in redox controlled expression in this organism.

Due to the proposed role of ADHE proteins in fermentative metabolism, is it not surprising that the expression of such proteins may be under regulation linked to the redox balance within the cell due to the accumulation of NADH. Experiments carried out during this project aimed to determine the expression pattern of various enzyme activities produced during fermentative metabolism.

1.5.6 Spirosome formation

ADHE proteins have been identified as forming large multimeric assemblies known as spirosomes (Bruchhaus and Tannich 1994; Espinosa 2001; Kessler et al. 1992). These large polymeric protein structures are made up of greater than 20 copies of the ADHE protein forming helical assemblies, as shown in Figure 1-6 A. The presence of these large protein assemblies was demonstrated through size exclusion chromatography, protein cross-linking and electron microscopy techniques by Kessler et al (1992). This work also demonstrated the dynamic nature of these structures, changing from an inactive "closed" form (without NAD^+ and Fe^{2+} present) to an "open" active form upon cofactor binding. This is shown in Figure 1-6 B, which is a schematic diagram of the ADHE conformational change.

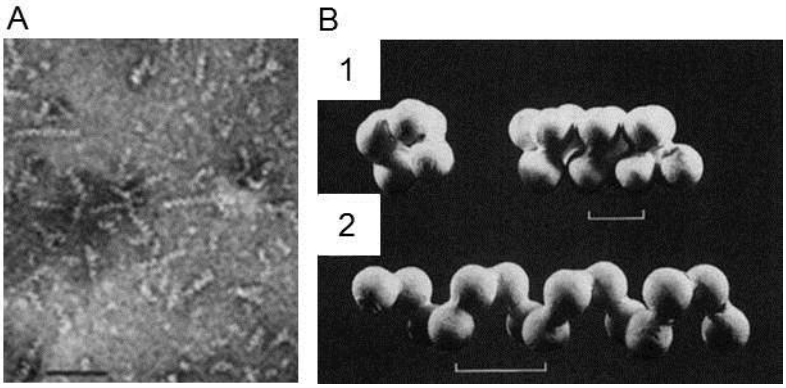


Figure 1-6: (A) Electron microscopy image of ADHE spirosomes (black line = 100 nm). (B) Schematic model of conformational change in spirosomes induced by the addition of 5 mM NAD^+, 0.3 mM Fe^{2+} and 5 µM CoA-SH. 1 = closed form, 2 = open (cofactor bound) form. Images adapted from Kessler et al (1992).

The reason for the formation of these complexes remains unclear, but if physiologically relevant they may provide some improvement of catalysis possibly through substrate channelling, or some stabilisation of the proteins may be conferred through assembly. The presence of such multimeric assemblies is investigated during this project for the TM242 ADHE protein.

1.5.7 Structural studies

No high-resolution structures for ADHE proteins are currently available in the protein data bank (PDB). One unpublished structure exists for the aldDH domain of the *Vibrio parahaemolyticus* ADHE (3MY7) to 2.25 Å resolution. No corresponding ADH domain structure is currently available. Structures of single domain proteins of a similar predicated fold exist for the two domains of ADHE. Although the overall architecture of such proteins is likely to be similar, the intervening loop regions may show significant variation. Potential interactions between the two domains of the protein due to their physical linkage may limit the validity of modelling work based on these homologous single domain proteins. Understanding the structural characteristics of the ADHE protein, especially in terms of the potential for substrate channelling, is of particular academic and industrial interest. Development of the structural understanding of the ADHE protein from TM242 has therefore become an objective of this project.

1.5.8 Pyruvate formate lyase inactivation

As well as functioning as both an aldDH and ADH, the ADHE protein from *E. coli* has also been described as having pyruvate formate lyase inactivating properties under aerobic conditions (Kessler et al. 1991; Kessler et al. 1992). This oxygen-sensitive enzyme is active through a “glycyl radical” which is stable anaerobically; upon switching to aerobic conditions the protein is cleaved at this catalytic residue and inactivated (Sawers and Watson 1998). ADHE is hypothesised to inactivate the enzyme prior to cleavage, allowing the enzyme to be re-activated upon switching to anaerobic conditions. Work carried out more recently has disputed this hypothesised inactivating activity (Nnyepi et al. 2007). As the pyruvate formate lyase has been knocked out of the TM242 strain, this debated activity of ADHE was not investigated during this project.

1.5.9 Ethanol tolerance

Inhibitory effects of alcohols on microorganisms in biofuel production are a commonly reported barrier to commercialisation (Liu and Qureshi 2009). There are many possible mechanisms by which alcohols such as ethanol inhibit the growth of microorganisms. These include effects on membrane integrity and cellular processes such as division and protein synthesis (Jones 1989). A commonly discussed cause is the effect on membrane integrity caused by increasing alcohol concentrations. Tolerant strains have been shown to vary the composition of the membrane fatty acids, possibly to overcome

increased membrane fluidity caused by increasing ethanol concentrations (Ingram et al. 1980; Jones 1989; Liu and Qureshi 2009; Timmons et al. 2009).

Unusually, the ADH domain of the ADHE protein in *Clostridium thermocellum* has recently been reported as an important factor in ethanol tolerance of this ethanol producing organism (Brown et al. 2011). Comparison of the genome sequence of a more ethanol tolerant strain to the parent strain showed many differences, although two point mutations in the ADH domain of the ADHE protein in isolation were demonstrated to influence the ethanol-tolerant phenotype of the organism. Specifically, a mutation of a conserved histidine residue (734) (the second in the conserved HSLAH motif shown in Figure 1-5) to an arginine was suggested to be responsible, the modified strain being able to tolerate higher concentrations of ethanol in the growth medium. This tolerant phenotype was not linked to increased levels of ethanol production. A reduction of 25-fold was observed in the specific activity of the ADH domain of ADHE compared to the wild type protein (assaying with NADH as the cofactor). An increase was observed in the specific activity of the enzyme with respect to NADPH as a cofactor (0.03 to 0.12 µg/min/mg). The relative increase in NADPH activity observed was low compared to the substantial loss of NADH dependent ADH activity (2.7 to less than 0.005 µg/min/mg). Ethanol production was not reported to drop significantly between the strains. Brown et al (2011) suggest that this cofactor switching may be linked to membrane composition, although this theory remains to be tested. A significant loss of activity (~70%) was observed in a mutant of *E. histolytica* where the same histidine residue (His_{734} in *E. histolytica*) was also mutated to an arginine. No effect in terms of ethanol tolerance was investigated (Espinosa 2001). It is difficult to rationalise the reported effect of these mutations on the phenotypic effect of ethanol tolerance. The significance of this work is further discussed during the structural investigation of the ADH domain of the TM242 ADHE protein described in Chapter 5 of this thesis.

1.6 Project aims and objectives

At the start of this project, an open aim was to characterise the ADHE protein from the *G. thermoglucosidasius* TM242 strain, which is used for the commercial production of bioethanol by TMO Renewables. The characterisation was to include kinetic and structural evaluation in an effort to develop an understanding of the properties of this thermophilic protein in terms of bioethanol production. Potential optimisation of the catalysis of acetyl-CoA to ethanol was of particular interest.

2 GENERAL MATERIALS AND METHODS

2.1 General laboratory reagents

In all cases, unless otherwise stated, laboratory reagents and bacteriological media were supplied by Sigma-Aldrich Company Ltd. (Poole, UK) or Fisher Scientific Ltd. (Loughborough, UK). DNA polymerases, restriction enzymes (BsaI, DpnI, HindIII, KpnI, NcoI, NdeI, NheI, PstI, SacI, XbaI & XhoI), dNTPs (deoxyribonucleoside triphosphates), T4 DNA ligase, DNA markers and associated reaction buffers were sourced from Promega (Southampton, UK) or New England Biolabs Ltd. (Hitchin, UK). Complete, EDTA-Free, Protease Inhibitor Cocktail Tablets and Shrimp Alkaline Phosphatase (SAP) were manufactured by Roche (Mannheim, Germany).

2.2 Strains, media and culture growth

2.2.1 *Escherichia coli (E. coli)*

E. coli BL21 (DE3) cells (Novagen): this strain is deficient in Lon and OmpT proteases to aid protein expression. *E. coli* BL21 (DE3) (Novagen) containing pTMO259 (pET28a) were supplied by TMO Renewables at the start of this study. This plasmid contains a kanamycin resistance gene as well as the gene of interest.

E. coli Rosetta© (DE3) competent cells (Novagen): this BL21 derivative strain contains the pRARE plasmid, which acts to overcome the difference in codon usage between *E. coli* and eukaryotic proteins (Novy 2001), and a gene coding for chloramphenicol resistance.

E. coli Arctic Express© (pRIL) (DE3) (Stratagene Texas USA): these cells contain the pRIL plasmid and plasmid-encoded cold-adapted chaperonin proteins. The pRIL plasmid acts to supply certain additional tRNA genes allowing expression of AT rich sequences. Chaperonins Cpn60 and co-chaperonin Cpn10 from psychrophilic bacterium *Oleispira antarctica,* are used to facilitate protein folding at low growth temperatures. These plasmids also convey streptomycin and gentamycin resistance. This strain is deficient in Lon and OmpT proteases (Stratagene 2009).

2.2.2 G. thermoglucosidasius

Geobacillus strains were supplied by TMO Renewables Ltd:

TM242 is the modified strain of *Geobacillus thermoglucosidasius* (*ldh*$^{-}$, *pfl*$^{-}$ *pdh*$^{+}$) discussed in the Introduction (Cripps et al. 2009).

TM444 is a modified version of the TM242 strain. The ability of the organism to sporulate has been removed through specific gene targeting.

TM89 is a modified strain of *Geobacillus thermoglucosidasius* where the lactate dehydrogenase gene has been removed (*ldh*$^-$).

TM400 is an ADHE knock-out strain produced from TM242 using a targeted integration system by TMO Renewables. A 336 bp (coding for 112 amino acids) section was removed from the middle of the *adhE* gene, covering sections of the predicted aldDH and ADH domains. This strain is unable to grow under anaerobic conditions.

TM393 is an ADHE knock-out strain produced from a TM89 parent strain using a targeted integration system by TMO Renewables (as in TM400). This strain is unable to grow under anaerobic conditions.

2.2.3 Cloning strains

E. coli JM109 (Promega) and DH5α (Invitrogen) strains were used for the preparation of plasmid DNA as these strains are *recA* negative, which ensures DNA stability. A β-galactosidase deficiency in these strains allows blue/white screening for the gene of interest being introduced via the plasmid pGEM®-T easy.

2.2.4 Bacterial growth media

Media were sterilised either by autoclaving or using 0.22 µm Steritop® filter units (Millipore) where required. Unless otherwise stated, solid plates of the different liquid media were made by adding 1.5% (w/v) agar prior to autoclaving.

Bacterial strains were cultured in liquid Lysogeny Broth (LB) medium (1% (w/v) NaCl, 1% (w/v) tryptone and 0.5% (w/v) yeast extract), or on solid plates (LB + 1.5% (w/v) agar), unless otherwise stated.

Other media types used during the project were:

Overnight Express Instant Terrific Broth (TB) medium: 6% (w/v) “Instant TB medium” (Novagen, Hull, UK) in 1L distilled water, supplemented with 1% (v/v) glycerol.

Yeast Extract Nutrient Broth (YENB) medium: 0.75% (w/v) yeast extract and 0.8% (w/v) nutrient broth (BD, NJ, USA).

Tryptone Soya Broth (TS): this medium was purchased pre-prepared from Oxoid, (Basingstoke, UK); the dehydrated medium was dissolved at 30 g/L in distilled water.

Super Optimal broth with Catabolite repression (SOC) medium: 2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose.

Soy Peptone Yeast extract (No Glucose) (2-SPY-NG) medium: 1.6% (w/v) soy peptone (Solabia, Paris, France), 1% yeast extract (Oxoid, Basingstoke, UK) and 0.5% (w/v) NaCl, adjusted to pH 7.0 with 5 M KOH.

Modified Two Tryptone Yeast extract (TTY) medium: 1.6% (w/v) Bacto-tryptone (BD, NJ, USA), 1% (w/v) Bacto-yeast extract (BD, NJ, USA) and 0.5% (w/v) NaCl.

50 mM Modified Urea Sulphate Medium (USM): 25 mM NaH_2PO_4, 50 mM urea, 25 mM K_2SO_4, 5 mM citric acid, 3.125 mM $MgSO_4$, 50 µM $CaCl_2$, 0.31 µM biotin, and 12.5 ml/L of trace elements, pH 7.0. Various concentrations of the appropriate sugar and yeast extract (Oxoid, Basingstoke, UK) were added as indicated.

The trace elements solution consisted of 60 mM H_2SO_4, 1.44 g/L $MnSO_4.7H_2O$, 5.56 g/L $FeSO_4.7H_2O$, 1.69 g/L $MnSO_4.H_2O$, 0.25 g/L $CuSO_4.5H_2O$, 0.562 g/L $CoSO_4$.7H2O, 0.06 g/L H_3BO_3 and 0.886 g/L $NiSO_4.6H_2O$ dissolved in MilliQ water.

Media were supplemented with antibiotics where required (chloramphenicol (chlor) 34 µg/ml, kanamycin (kan) 30 µg/ml (12 µg/ml for *Geobacillus* pUCG18 carrying strains), gentamycin (gent) 20 µg/ml, streptomycin (strep) 100 µg/ml or carbenicillin (carb) 50 µg/ml.

2.2.5 Strain storage

Strains used during this project were stored long-term as glycerol stocks at -80°C in cryogenic vials. Aliquots of 10 ml overnight cultures in the required growth medium were used for creation of these stocks. *E. coli* strains were stored as 80% (v/v) overnight culture: 20% (v/v) sterile glycerol, whilst *G. thermoglucosidasius* strains were stored as 70% (v/v) overnight culture: 30% (50% v/v) sterile glycerol (diluted in sterile MilliQ H_2O). These stocks were mixed thoroughly and snap-frozen in liquid nitrogen prior to immediate transfer to storage at -80°C. Unless otherwise stated, a loop-full of frozen glycerol stock was used to inoculate 10 ml overnight cultures used during

experiments performed here. Agar plates of the desired growth medium were used for short-term storage of strains at 4°C where required. Single colonies were used to inoculate overnight cultures.

2.3 Molecular biology

Prior to this study, the whole length *adhE* gene of interest was cloned into the expression vector pET28a (Novagen), which carries kanamycin resistance, by Novacta on behalf of TMO Renewables. The gene was introduced between the NheI/XhoI restriction sites of the vector. This plasmid is known as pTMO259 and was transformed into the various expression strains used in this study. Isolated pTMO259 and TM242 genomic DNA was supplied by TMO Renewables.

Vector maps for the various plasmids used during this project are included in Appendix 1, as are the gene sequences and corresponding protein amino acid sequences for the gene products investigated here.

The amino acid sequence of the ADHE protein was determined from the *adhE* gene sequence annotated in the TM242 genome using the Translate tool (ExPASy (Gasteiger E. 2005)).

2.3.1 Primer design

DNA primers were manufactured by Invitrogen (Paisley, UK), and a complete list of primers used during this project can be seen in Appendix 1. Forward primers were designed to be identical to the desired 5' region of the gene of interest, and reverse primers to the desired 3' end of the target gene. Approximately 20 bp of coding sequence (dependent on the desired T_m of the primers) was used. Where required, DNA primers incorporated an appropriate restriction site at the desired terminus of the DNA sequence, along with stop codons in the reverse primers. Restriction enzymes selected for use during this project were selected for their inability to cut the particular gene sequence of interest, along with the presence of the required restriction site within the cloning/expression vector to be used. As recommended by the manufacturers, some additional non-coding DNA (~1-6 bp) was sometimes added, flanking the restriction site in the primers to enhance cleavage using certain restriction enzymes.

2.3.2 Polymerase chain reaction (PCR)

PCRs using Phusion® DNA polymerase (Finnzymes, Finland) were carried out in an Eppendorf Mastercycler (Eppendorf, Germany) as described in the manufacturer's

instructions. Unless otherwise stated, the reactions used purified plasmid DNA as template. DNA primers were designed to introduce desired restriction sites at the appropriate termini where required for subsequent digestion. A list of DNA primers used during this project is included in Appendix 1. Primers were diluted in nuclease-free MilliQ H_2O as required. The PCR protocol used during this project is summarised in Table 2-1 and Table 2-2.

Reagent	Volume added (to 20 µl reaction)	Final concentration
Nuclease-free MilliQ H_2O	Adjusted to final volume of 20 µl	
5x Phusion® HF buffer	4 µl	1x
Primer 1 (forward) (5 µM)	1 µl	0.25 µM
Primer 2 (reverse) (5 µM)	1 µl	0.25 µM
dNTPs (10 mM)	0.4 µl	200 µM
DNA template	0.5 µl	Variable
DNA polymerase (Phusion®)	0.2 µl	0.4 U

Table 2-1: Overview of standard PCR reagents used during this project.

Step	Temperature (°C)	Time (s)
Initial denaturation	98	30
Denaturation	98	10
Annealing	~3 below primer $T_{melting}$	20
Extension	72	30 (per kb of target DNA sequence)
Final extension	72	600

Table 2-2: Overview of standard PCR protocol used during this project.

The denaturation-extension steps were cycled 30-35 times to allow sufficient target DNA to be amplified.

For PCR colony screening of transformants, a *Taq* polymerase Master Mix was used (Genesys, UK) according to the manufacturer's instructions. A small amount of bacterial colony was used to provide the DNA template required to screen using the desired primers. The programme used was as described in Table 2-2 with the following adjustments: a 40 s 95°C heat lysis step was included at the start of the programme. Denaturation steps were carried out at 95°C, annealing was at 55°C and extension time was 1 min/kb target gene. The PCR reaction mixture was run on an agarose gel to identify positive colonies.

2.3.3 Site-directed mutagenesis

Site-directed mutagenesis was used to make small changes to DNA coding regions of various cloned genes produced during this project. Mutagenesis primers were designed using the QuikChange® primer design program (Agilent technologies CA, USA) and mutagenesis reactions were carried out according to the QuikChange®II site-directed mutagenesis kit protocol. The methylated and hemi-methylated template DNA was digested using DpnI prior to transformation of the mutated plasmid. A list of DNA primers used in this project is included in Appendix 1.

2.3.4 Agarose gel electrophoresis

Electrophoresis in 1% (w/v) agarose gels supplemented with 0.5 µg/ml ethidium bromide, produced by dissolving in TAE buffer (40 mM Tris (2-Amino-2-hydroxymethyl-propane-1,3-diol) acetate (pH 8.0), 1 mM EDTA), was used for DNA visualisation. DNA was prepared by dilution with 6x loading dye (50% (v/v) glycerol, 50 mM EDTA, pH 8.0, 0.05% (w/v) bromophenol blue). TAE buffer was used as electrophoresis tank buffer and DNA was run at a constant voltage (between 70 and 80V). The electrophoresis was monitored by following the dye front, and the gels were run until the desired band separation was obtained. DNA-ethidium bromide was visualised using a UV transilluminator. Co-electrophoresis of samples with standard DNA markers was used to estimate the size of DNA fragments.

2.3.5 DNA purification

DNA bands were excised from agarose gels with clean razor blades and then purified using the Promega Wizard SV gel and PCR clean-up system following the manufacturer's protocol (Promega 2010b). An elution volume of 30-50 µl MilliQ water was used and DNA was stored at -20°C.

2.3.6 A-tailing

To allow efficient ligation into the pGEM®-T easy vector, gel-purified blunt-ended PCR fragments were A-tailed at the 3' ends using *Taq* DNA polymerase, following the Promega A-tailing protocol (pGEM®-T Vector systems manual, 11th edition, 2006). The reaction was carried out at 70°C for 30 min prior to purification using the Promega Wizard SV gel and PCR clean-up system. This incorporated a single 3'-adenine at both ends of the PCR fragment that is complementary to the terminal thymines present in the vector.

2.3.7 pGEM®-T easy

The cloning vector pGEM®-T easy was used to amplify high yields of the target genes produced during this project, prior to restriction digestion with appropriate enzymes and ligation into the protein expression vectors used. A-tailed PCR fragments were ligated into the linearized vector using the overhanging 3' terminal thymine present at both ends of the vector. The use of this vector also allowed convenient screening of transformants using blue/white screening methods.

2.3.8 Ligation

Target gene inserts were ligated into cut plasmid vectors using the Promega rapid ligation kit, according to the protocol provided (using the supplied 10x ligation buffer). The 10 µl ligation reactions used a range of ratios of insert to vector and were carried out at 16°C overnight.

2.3.9 Ethanol precipitation

To purify DNA from ligations of DNA inserts and cloning/expression vectors, the ligation mixture was ethanol precipitated ready for electroporation using the following protocol:

- 20 µl MilliQ H_2O was added to the ligation, which was then heat treated for 20 min at 70°C.
- The mixture was cooled on ice prior to the addition of 3.1 µl of 2.5 M sodium acetate (pH 5.2), 1 µl of dextran blue (0.15 g dextran blue, 0.5 ml 0.5 M EDTA pH 8.0, 9.5 ml MilliQ H_2O) and 68.2 µl of cold (-20°C) ethanol.
- The solution was incubated at -20°C for 30 min prior to centrifuging for 20 min at 16,000 x *g* (4°C).
- The supernatant was removed and the pellet washed with 100 µl of 70% (v/v) ethanol.
- The solution was centrifuged for 10 min at 16,000 x *g* (4°C).
- The supernatant was removed and the pellet washed with 150 µl of 70% (v/v) ethanol.
- The solution was centrifuged for 10 min at 16,000 x *g* (4°C) and the supernatant was removed.
- The pellet was allowed to air-dry prior to being resuspended in 5 µl of nuclease-free MilliQ H_2O.

2.3.10 Blue/white colony screening

Initial screening of JM109/DH5α transformants of the pGEM®-T easy cloning vector was carried out using blue/white screening. Ligation of the target gene sequence into the multiple cloning site of the pGEM®-T easy cloning vector results in the disruption of the *lacZα* gene. This prevents the restoration of β-galactosidase activity in transformed strains. The phenotype of conversion of X-gal (5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside) to a blue insoluble precipitate is not observed for transformants carrying the successfully ligated pGEM®-T easy-target gene constructs. White colonies were selected for screening using other methods to confirm successful transformation and ligation of the pGEM®-T easy-target gene constructs.

Cells were spread onto LB-IPTG-X-Gal agar plates supplemented with carbenicillin (previously spread with 100 µl of 100 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) and 20 µl of 50 mg/ml X-Gal (Melford) dried for 30 min) and incubated overnight at 37°C.

2.3.11 Plasmid DNA preparation

Cultures of plasmid-carrying strains were grown overnight at 37°C in 10 ml of LB (in a 50 ml capacity falcon tube) with appropriate antibiotic selection and shaking at 200 rpm, and were harvested by centrifugation at 5,300 x *g* for 10 min. Plasmids were subsequently purified using the Promega Wizard plus SV minipreps DNA purification system, according to the supplied protocol (Promega 2010a). DNA was eluted in 30-50 µl MilliQ water and was stored at -20°C.

2.3.12 Restriction digests

5 µl of purified plasmid DNA was used for restriction digests (total volume 20 µl made up with nuclease free MilliQ H_2O). Where required, BSA (bovine serum albumin) at a concentration of 100 µg/ml was added to the reactions. For digests with two restriction enzymes, a reaction buffer that was suitable for both enzymes was selected. Approximately 5 U of the appropriate restriction enzyme was added to the reaction (star activity was avoided by keeping glycerol concentrations below 10% v/v) and the reaction was incubated at 37°C for 2-3 h. Vector digestion reactions were halted through heating at 65°C, prior to immediate treatment with SAP (10 U at 37°C for 1 h). This procedure dephosphorylated the cut DNA post digestion, and so prevented re-ligation in the absence of an insert. Following incubation, digested samples were separated immediately on agarose gels and purified as described in Section 2.3.5.

2.3.13 DNA sequencing

To confirm successful cloning of the target gene sequences described in this project into the required cloning/expression vectors, plasmid DNA was sent for sequencing by Source BioScience, UK. The genes were sequenced with vector or gene specific primers, including internal primers where required.

2.3.14 Preparation of electrocompetent *E. coli*

Unless otherwise stated, electrocompetent *E. coli* strains were prepared in-house using the following protocol. *E. coli* strains were grown in YENB media in baffled conical flasks until mid-exponential phase (inoculated from a 10 ml overnight culture). Cells were harvested by centrifugation at 4,500 x *g* for 10 min at 4°C. All subsequent steps were carried out on ice and using chilled equipment and solutions. Cells were resuspended and washed with MilliQ H_2O prior to harvesting by centrifugation; this was repeated 3 times, and the final resuspension was in 10% (v/v) sterile glycerol in MilliQ H_2O. Aliquots (30-50 µl) of the cell solution were snap-frozen in dry ice and stored at -80°C until required.

2.3.15 Electroporation (transformation) of electrocompetent *E. coli* strains

To transform electrocompetent strains of *E. coli* with the desired plasmid, the following protocol was followed. Aliquots of cells (30-50 µl) were removed from storage at -80°C and left to thaw on ice for 10 min. 0.1 cm gapped electroporation cuvettes (BioRad) were placed on ice at this stage, along with the plasmid preparation to be transformed. 0.5 µl of plasmid preparation was inoculated into the thawed cell aliquots and incubated on ice for 2 min. The cell aliquots were then transferred to the chilled cuvettes, placed in the micropulser electroporator (BioRad) and subjected to a 5 m-s electrical pulse of approximately 1.8 kV. The cells were resuspended in 1 ml of SOC medium and grown with shaking at 37°C for 1 h without selection. Between 10 and 200 µl of cells were then plated onto LB agar plates containing the appropriate selection antibiotics, and grown at 37°C overnight.

2.3.16 Transformation of chemically-competent *E. coli* strains

The commercially-produced, chemically-competent *E. coli* BL21 (DE3) cells (Novagen), were used during this project. Transformation used the following procedure. Aliquots of cells (20 µl) were removed from storage at -80°C and left to thaw on ice for 10 min. 1 µl of plasmid preparation was inoculated into the thawed cell aliquots and incubated on ice for 20 min. The cell aliquots were then incubated at 42°C for 45 s, following which the cells were immediately cooled on ice for a minimum of 3 min. The cells were

resuspended in 1 ml of SOC medium and grown with shaking at 37°C for 1 h without selection. Between 10 and 200 µl of cells were then plated onto LB agar plates containing the appropriate selection antibiotics, and grown at 37°C overnight.

2.3.17 Preparation of electrocompetent *Geobacillus* sp

Electrocompetent *Geobacillus* strains were prepared in-house using the following protocol modified from Cripps et al (2009). *Geobacillus* strains were grown in pre-warmed TTY media in baffled conical flasks (inoculated from a 10 ml overnight culture) until early-exponential phase (OD_{600nm} 1.4-2) at 55°C with shaking at 225 rpm. Cells were harvested by centrifugation at 3,220 x *g* for 20 min at 4°C. All subsequent steps were carried out on ice and using chilled equipment and solutions. Cells were resuspended and washed with filter-sterilised electroporation medium (0.5 M sorbitol, 0.5 M mannitol, 10% glycerol) prior to harvesting by centrifugation; this was repeated 3 times, and the final resuspension volume was adjusted to give a final concentration of approximately 40 OD_{600nm}/ml. 60 µl aliquots of the cell solution were stored at -80°C.

2.3.18 Electroporation (transformation) of electrocompetent *Geobacillus* strains

To transform electrocompetent strains of *Geobacillus* with the desired plasmid, the following protocol, modified from Cripps et al (2009), was followed. Aliquots of cells (60 µl) were taken from storage at -80°C and left to thaw on ice for 10 min. 0.1 cm gapped electroporation cuvettes (BioRad) were placed on ice at this stage, along with the plasmid preparation to be transformed. 2 µl of plasmid preparation was inoculated into the thawed cell aliquots and incubated on ice for 5 min. The cell aliquots were then transferred to the chilled cuvettes, placed in the micropulser electroporator (BioRad) and subjected to a 5 m-s electrical pulse of approximately 2.5 kV. The cells were resuspended in 1 ml of pre-warmed (55°C) TTY medium and grown with shaking at 200 rpm and 55°C for 2 h without selection. 200 µl of cells were then plated onto TTY agar plates (pre-warmed at 55°C) containing the appropriate selection antibiotic, and grown at 55°C overnight.

2.4 Protein expression

Unless otherwise stated, *E. coli* protein expression cultures were prepared by the following method. A 10 ml starter culture in LB medium was grown in the presence of selection antibiotics (if appropriate) with shaking at 250 rpm, 37°C, overnight. This culture was used to inoculate the larger-scale expression culture (1 in 100). This culture was grown at the required growth temperature (see appropriate Results sections) until the OD_{600nm} reached the range between 0.5 and 1.0, at which point any

necessary temperature change was made and the culture left to equilibrate (if required) for approximately 25 min, prior to the culture being supplemented with 1 mM IPTG for induction.

Protein expression vectors commonly used during this project allowed modulation of protein expression via a T7 promoter. Expression of the target gene is normally repressed by the lac repressor protein. IPTG is an analogue of allolactose which binds the lac repressor protein, alleviates repression and thus induces protein expression (Novagen 2011).

2.5 *Geobacillus* culturing techniques

2.5.1 Tube fermentation method

This method of tube fermentation was developed by TMO Renewables Ltd.

- TSA plates inoculated with 150 µl of thawed glycerol stock were prepared for each strain to be investigated; these plates were incubated at 60°C for approximately 16 h prior to the tube fermentation experiment.
- Seed tubes were prepared by inoculating 5 ml of pre-warmed (60°C) 2-SPY-NG medium in 50 ml falcon tubes with a loop-full of biomass from the TSA plates. These tubes were then incubated at 60°C with shaking at 240 rpm for 6 h.
- 1 ml of this seed culture was then used to inoculate 10 ml of pre-warmed (60°C) USM medium (2% glucose 0.5% yeast extract) in 15 ml falcon tubes. These tubes were then incubated at 60°C with shaking at 240 rpm for 42-48 h to allow tube fermentation to take place.
- The sample tubes were then centrifuged at 3,220 x *g* for 20 min at 4°C and the supernatant was transferred into fresh tubes prior to analysis.
- Metabolite concentrations were analysed at TMO Renewables using High Performance Liquid Chromatography (HPLC).

2.5.2 Overnight shake-flask method

This method was used to produce aerobically-grown *Geobacillus* strains within the exponential phase of growth. This protocol is a modified version of a method developed by TMO Renewables Ltd.

- A 50 ml falcon tube containing 25 ml of the desired medium (pre-warmed to 60°C) was inoculated with 50 µl of thawed glycerol stock of the desired strain.

- 1 ml of this dilution was used to inoculate 50 ml of medium in a 250 ml baffled flask (pre-warmed to 60°C).
- The cultures were incubated for 16 h at 60°C with shaking at 220 rpm.

Cell pellets produced by this method were obtained by centrifugation at 5,300 x *g* for 20 min (4°C). The supernatant was discarded and the pellets stored at -20°C until required.

2.6 Cell extract preparation

Cells were harvested by centrifugation at 5,300 x *g* for 20 min (4°C). The supernatant was discarded and the pellets were resuspended in one-tenth of the original culture volume of HIS-BIND buffer (50 mM Tris buffer, 300 mM NaCl, pH 8.0, and 20 mM imidazole (Acros Organics)) unless otherwise stated. Protease-inhibitor tablets (Roche) were added at a minimum concentration of 1 tablet to 10 ml sample volume, as per the manufacturer's instructions. Cells were then sonicated on ice at 14 microns amplitude; small culture preparations (~2 ml) were given 3 x 10 s bursts with 15 s incubation on ice between each burst, whereas large preparations (10 ml maximum) were given 5 x 20 s bursts with 20 s incubation on ice between each burst. Further centrifugation (16,000 x *g* for 18 min at 4°C) allowed separation of the soluble and insoluble fractions. The supernatant (soluble protein fraction) was isolated, and the pellet (insoluble fraction) was resuspended in a one-tenth volume of the original culture of re-suspension buffer (8 M urea).

Geobacillus cell extracts were prepared in a similar manner in 50 mM EPPS buffer, pH 8.0 (0.3 g cells/ml buffer) with 1 mM AEBSF (4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride); 1.5 µl of Benzonase© nuclease (250 U/µl) (Sigma) per ml of resuspended cells was also added.

2.7 Sodium Dodecyl Sulphate PolyAcrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE analysis was performed using a Mini-PROTEAN electrophoresis system (BioRad). Gels (10%) were produced in-house using the components described in Table 2-3. Ammonium Per-Sulphate (APS) and Tetramethylethylenediamine (TEMED) were added immediately prior to casting to catalyse polymerisation of the gel. Gels were run in Tris/glycine buffer, pH 8.3 (10x stock = 30 g Tris, 144 g glycine and 100 ml of 10% (w/v) SDS made up to 1L with MilliQ H_2O).

Resolving gel		Stacking gel	
Component	**Volume added**	**Component**	**Volume added**
MilliQ H_2O	4.0 ml	MilliQ H_2O	2.2 ml
1.5 M Tris/HCl pH 8.8	2.5 ml	0.5 M Tris/HCl pH 6.8	1.0 ml
30% (w/v) acrylamide	3.3 ml	30% (w/v) acrylamide	0.66 ml
10% (w/v) SDS	100 µl	10% (w/v) SDS	40 µl
10% (w/v) APS	100 µl	10% (w/v) APS	40 µl
TEMED	10 µl	TEMED	10 µl

Table 2-3: SDS-PAGE gel components.

Where required, samples were diluted with the appropriate buffer prior to diluting 1:4 with 4x SDS-PAGE loading buffer (stock: 4 ml 0.5 M Tris/HCl, pH 6.8, 2 ml 10% (w/v) SDS, 4 ml glycerol, 100 µl β-mercaptoethanol and 8 mg bromophenol blue) and denatured by heating at 95°C for 3 min prior to loading onto the SDS-PAGE gel (10 µl/lane). Gels were electrophoresed at 80V until the sample had passed through the stacking gel, after which the voltage was increased to 200V. The gels were subsequently stained with Coomassie brilliant blue R250, de-stained over 4 h using 3-4 changes of a 60:30:10 distilled water: methanol: glacial acetic acid solution, and subsequently dried.

Broad-range M_r markers (Bio-Rad) (6 µl) were used to estimate the M_r values of the proteins by co-electrophoresis with the protein samples. The log M_r values of the markers were plotted against their R_f values (distance travelled by the protein band/distance travelled by gel front) and a straight line of best fit applied. The R_f value of the protein of interest was then used to estimate the M_r.

The predicted M_r of the various proteins investigated in this project were calculated using the ProtParam tool from the ExPASy website (Gasteiger E. 2005).

2.8 Estimation of protein concentration

Where possible, the concentration of protein present in purified protein samples was determined using the absorbance at 280 nm (A_{280nm}), using absorption coefficients determined for the protein of interest using the ProtParam tool.

Where the purity of the protein sample analysed was not high enough to use the method described above, an estimate of protein concentration was made using the Bradford assay (Bradford 1976). This used the Bio-Rad protein assay reagent

(Coomassie brilliant blue G250) comparing samples to BSA standards (between 0 and 10 µg/ml).

2.9 Metal-affinity purification

Some of the vectors used for expression of the recombinant proteins allowed the incorporation of a histidine "Tag" at the N- or C-terminal end of the protein (depending on the restriction sites used). Metal-affinity columns use the affinity of poly-histidine tags for charged Ni^{2+} to separate the proteins of interest from other proteins produced by the expression strain. The column was prepared according to the manufacturer's instructions and the resin (metal-chelating cellulose) (Bioline) was charged with $NiSO_4 \cdot 6H_2O$ (2 ml of a 400 mM solution). The column was washed with distilled water and equilibrated with HIS-BIND buffer prior to loading the soluble protein fraction onto the column. The various fractions were collected in separate tubes and stored on ice.

The initial fraction eluted was reloaded to ensure that all tagged protein became bound to the column. Following this, the column was washed with 5 column volumes of HIS-BIND buffer to elute any unbound material. An increasing concentration of HIS-ELUTE buffer (50 mM Tris buffer pH 8.0, 300 mM NaCl and 1 M imidazole, diluted with HIS-BIND buffer where required) was used to elute bound proteins stepwise in order of affinity (5 ml elution for each step). These elution solutions were adjusted where required to maximise the purity of the protein obtained. Following collection of the fractions, samples were taken for SDS-PAGE to confirm the presence and purity of the protein product obtained.

Occasionally, a 5 ml His-Trap HP column on an ÄKTA Explorer system (GE Healthcare) was used for purification of His-tagged proteins, using a continuous gradient of imidazole as eluate.

2.10 Dialysis

Dialysis tubing of the appropriate size was prepared by boiling for 10 min in 2% (w/v) sodium bicarbonate pH 8.0, and 1 mM EDTA, followed by 10 min boiling in 1 mM EDTA pH 8.0, alone. Rinsing with distilled water followed each treatment. Protein samples were sealed within the tubing and placed in a large volume (normally 2 l) of the appropriate dialysis buffer and dialysed at 4°C with slow stirring overnight.

2.11 Ion-exchange chromatography

Ion-exchange chromatography was carried out at 25°C using 5 ml HiTrap SP HP (Cation) or HiTrap Q HP (anion) columns on an ÄKTA Explorer Fast Protein Liquid Chromatography (FPLC) system (GE Healthcare). The buffers selected were at least 1 pH unit higher/lower (Anion/Cation) than the predicted pI of the target protein; columns were pre-equilibrated with running buffer prior to loading of the sample. A gradient of 0-1 M NaCl was used to elute the protein. The elution profile of the protein was detected by A_{280nm} and those fractions over which the protein eluted were collected and stored on ice.

2.12 Gel filtration

Unless otherwise stated, gel filtration was carried out at 25°C on a Superdex 200 10/300GL gel filtration column, on an ÄKTA Explorer FPLC system (GE Healthcare). The column was calibrated using various M_r standards (Ribonuclease M_r 13,700, Ovalbumin M_r 43,000, Conalbumin M_r 75,000, Aldolase M_r 158,000 and Ferritin M_r 440,000) (GE Healthcare). The markers selected were dependent on the M_r of protein sample run. The plot used to estimate the M_r of the protein of interest was Log M_r vs K_{av} where K_{av} = (Ve-Vo)/(Vc-Vo) (Ve = elution volume, Vo = Column void volume, and Vc = geometric column volume).

Where required, protein samples were filtered using an Ultrafree centrifugal filter device (0.45 µm) (Millipore) and concentrated using a Vivaspin 5K MWCO centrifugal filter device (Sartorius), prior to loading onto the column following the manufacturer's instructions (pre-rinsing with MilliQ H_2O).

The elution profile of the protein was detected by A_{280nm} and those fractions over which the protein eluted were collected and stored on ice.

2.13 Enzyme activity assays

The ADHE protein studied in this project is thought to possess two separate enzymatic activities, both of which were measured spectrophotometrically at 60°C (the optimum growth temperature of the native organism). A Varian Cary 50 Bio UV/visible light spectrophotometer with a Peltier temperature controller was used for enzyme assays. Specific activities (µmol/min/mg) were determined using the molar absorption coefficient of the monitored substrate, and the concentration of protein determined for the enzyme sample. Substrate-independent background rates were measured for all assays and taken into account in the determination of the final rate of reaction.

Substrates were dissolved and diluted in the required assay buffer whenever possible. The assays described below describe the standard assay mixture used for assays throughout this project. Variations on these assays will be discussed alongside the results. In all assays, it was shown that the rate of reaction was directly proportional to the enzyme concentration.

2.13.1 Acetaldehyde dehydrogenase (aldDH)

$$\text{Acetyl-CoA} + \text{NADH} + H^+ \rightleftharpoons NAD^+ + \text{acetaldehyde} + \text{CoA-SH}$$

Two assay methods were developed for this enzyme activity, one measuring the appearance of CoA-SH with DTNB (5,5′-dithiobis-(2-nitrobenzoic acid)), and the second directly monitoring the disappearance of NADH.

2.13.1.1 AldDH DTNB assay

DTNB is cleaved by thiol compounds (e.g. CoA-SH) stoichiometrically to release a yellow-coloured compound (2-nitro-5-thiobenzoate^{2-}) (NTB) (Ellman 1959). This compound can be quantified using its absorbance of light at 412 nm using a spectrophotometer (ε_{412} = 13,600 $M^{-1}.cm^{-1}$), allowing the progress of the reaction to be monitored over time (CoA-SH release). This assay is only suitable for reactions carried out at neutral or alkaline pH as the yellow colour comes from the anion^{2-}, which only forms within the neutral-alkaline pH range. The components used for the assay are described in Table 2-4.

Chemical	Stock solution concentration (mM)	Cuvette concentration (mM)	Volume added to cuvette (ml)
Buffer	-	-	0.91*
NADH	~4**	~0.24	0.06
Acetyl-CoA	~7**	~0.14	0.02
DTNB	10	0.1	0.01

Table 2-4: aldDH DTNB assay components.

- *Volume of buffer and protein sample should add up to 0.91 ml
- **Each batch of acetyl-CoA and NADH was quantified prior to use
- Buffer pre-incubated at 60°C
- Substrates stored on ice
- Reaction started with protein sample addition

2.13.1.2 AldDH NADH assay

NADH has a distinct absorbance peak at 340 nm which is not present for the oxidised form (NAD^+); this allows the oxidation reaction to be monitored directly by the

acetyl-CoA dependent loss of absorbance over time at 340 nm (NADH) (ε_{340} = 6,220 $M^{-1}.cm^{-1}$). This assay was used when activity determination at acidic pH was required, rendering the DTNB assay impossible. The components used for the assay are described in Table 2-5.

Chemical	Stock solution concentration (mM)	Cuvette concentration (mM)	Volume added to cuvette (ml)
50 mM Citric acid pH 6.0 0.1 mM zinc acetate	50	~50	0.92*
NADH	~4**	~0.24	0.06
Acetyl-CoA	~7**	~0.14	0.02

Table 2-5: aldDH NADH assay components.

- *Volume of buffer and protein sample should add up to 0.92 ml
- **Each batch of acetyl-CoA and NADH was quantified prior to use
- Buffer pre-incubated at 60°C
- Substrates stored on ice
- Reaction started with protein sample addition

2.13.2 Alcohol dehydrogenase (ADH) (Acetaldehyde reduction)

$$\text{Acetaldehyde} + \text{NADH} + H^+ \rightleftharpoons NAD^+ + \text{ethanol}$$

This assay monitored the acetaldehyde-dependent loss of absorbance over time at 340 nm (NADH) (ε_{340} = 6,220 $M^{-1}.cm^{-1}$). The components used for the assay are described in Table 2-6.

Chemical	Stock solution concentration (mM)	Cuvette concentration (mM)	Volume added to cuvette (ml)
50 mM Citric acid pH 6.0 0.1 mM zinc acetate	50	~50	0.89*
NADH	~4**	~0.24	0.06
Acetaldehyde	4000	200	0.05

Table 2-6: ADH NADH assay components.

- *Volume of buffer and protein sample should add up to 0.89 ml
- **Each batch of NADH was quantified prior to use
- Buffer pre-incubated at 60°C
- Substrates stored on ice
- Reaction started with protein sample addition

2.13.3 Acyl-CoA preparation and quantification

Acyl-CoAs were prepared by the acylation of coenzyme A (CoA-SH) with the required acyl-anhydride using the following protocol: a solution of 10 mg of coenzyme A dissolved in 1 ml of deionised water was cooled on ice for 10 min prior to the addition of 0.2 ml of 1 M potassium bicarbonate followed by 5 µl of the required acyl-anhydride. This reaction mixture was left for at least 10 min on ice to allow complete acylation to occur. The extent of acylation was checked by monitoring the reaction between the acyl-CoA solution and DTNB at 412 nm to ensure significant levels of free coenzyme A were not present.

The stock concentration of acetyl-CoA was quantified using pig heart citrate synthase. This enzyme catalyses the following reaction, which goes to >99% completion (Srere 1975).

$$\text{Acetyl-CoA} + \text{Oxaloacetate} + H_2O \rightarrow \text{Citrate} + \text{CoA-SH}$$

Chemical	Stock solution concentration (mM)	Cuvette concentration (mM)	Volume added to cuvette (ml)
20 mM Tris/HCl pH 8.0	20	~20	0.96*
Oxaloacetate	20	0.20	0.01
DTNB	10	0.20	0.02
Acetyl-CoA	TBC	TBC	0.01
Citrate synthase	8.7 mg/ml (≥100 U mg^{-1})	0.87 mg	0.001

Table 2-7: Acetyl CoA quantification assay components.

- *Volume of buffer and citrate synthase should add up to 0.96 ml
- Reaction carried out at room temperature

The spectrophotometer was calibrated with the above reaction mixture prior to the addition of excess citrate synthase. The reaction was allowed to progress to completion prior to measuring the absorbance at 412 nm, allowing the stock concentration of acetyl-CoA to be calculated via the concentration of NTB^{2-} anion produced (ε_{412} = 13,600 $M^{-1}.cm^{-1}$). As pig heart citrate synthase will not catalyse the reaction with other acyl-CoAs, it was assumed that comparable concentrations of the other acyl-CoAs were produced by this method.

2.13.4 NAD^+ : CoA released ratio assays

As the product of the aldDH domain is the substrate for the ADH domain of ADHE, assays were designed to measure the metabolic flux to ethanol which may be

observed within assays. The ratio of NADH converted relative to CoA-SH released was determined using the following method.

These assays consisted of two parts:

1) The aldDH NADH assay was carried out as in 2.13.1 for 30 s at 60°C. The change in A_{340nm} was measured.
2) The aldDH NADH assay was carried out as above for 30 s at 60°C; at 30 s 100 µl DTNB (1 M Tris pH 8.8) was added to shift the pH to pH 8.5 and the change in A_{412nm} was measured.

Unless otherwise stated, each of these assays was carried out in triplicate and the results were used to calculate the ratio of NADH converted relative to CoA-SH released using the absorption coefficients of NADH (ε_{340} = 6,220 $M^{-1}.cm^{-1}$) and NTB^{2-} (ε_{412} = 13,600 $M^{-1}.cm^{-1}$).

NADH oxidase (acetyl-CoA free) and PAT (NADH free) background rates were measured and taken into account in the calculation of the final result of these assays.

2.13.5 Optimum temperature (T_{opt}) assays

Assays to determine the optimum temperature of various enzyme activities were carried out during this project. The assay buffer and spectrophotometer were pre-heated to the desired temperature prior to the assay. The temperature of the reaction mixture immediately following the assay was recorded as the actual temperature of the assay. The results were plotted as % activity of the maximum observed against temperature. These assays were carried out in citric acid buffer, pH 6.0, 0.1 mM zinc acetate; the pH of this buffer does not fluctuate with temperature and thus could be used across the temperature range without adjustment.

2.13.6 Thermostability (T_{stab}) assays

Assays to determine the stability of various enzyme activities at a range of temperatures were carried out during this project. The enzyme solution was aliquoted into a number of 0.5 ml Eppendorf tubes (not more than 100 µl). One tube was kept on ice for the duration of the experiment to act as a zero time point. The other tubes were incubated in a pre-heated heat block at the desired temperature for a given period of time, at which point the tube was removed and placed immediately on ice. A pre-heated metal block (100°C) was placed on top of the lids of the tubes during the experiment to prevent evaporation. Any precipitated protein was removed by

centrifugation at 14,000 x *g* (4°C) for 5 min. Standard assays were carried out on the heated samples to determine the level of activity remaining following incubation. The results were plotted as % activity and ln % activity of the maximum observed against time of incubation at the specified temperature.

2.13.7 Kinetic parameter determination

Where enzyme activities obeyed Michaelis-Menten kinetics, kinetic parameters were determined using the direct linear method (Eisenthal and Cornish-Bowden 1974) using the EnzPack software programme (Biosoft). Data are displayed as Michaelis-Menten plots and Hanes fits.

2.13.7.1 Standard error

Standard errors of the kinetic parameters obtained were calculated using the 68% confidence limits associated with the direct linear determination of kinetic parameters as outlined below:

$$Standard\ Error\ = \frac{SD}{\sqrt{(n-1)}}$$

(where the 68% confidence limits encompasses two SDs and n is the number of data points)

2.13.7.2 Percentage saturation

With the two substrate reactions, it was not always possible to saturate the enzymes with the fixed substrate when determining kinetic parameters. For example, the maximum concentration of NADH which could be used for these assays was 0.24 mM (A_{340nm} = 1.5). Also due to the reactive nature of acetaldehyde it was decided that the concentration of this substrate should not exceed 400 mM in the assays. Therefore where possible, the V_{max} values obtained through the assays under these non-saturating conditions were adjusted, based upon the predicted % saturation of the substrates involved, calculated through rearrangement of the Michaelis-Menten equation. The errors associated with this value were also accounted for. An example of this calculation is shown below:

The Michaelis-Menten equation: $v = \frac{Vmax\,[S]}{Km+[S]}$

Where v = rate of reaction, V_{max} = maximum theoretical rate of reaction, [S] = concentration of substrate and K_m = the Michaelis constant (concentration of substrate required to reach ½ V_{max}).

Rearranging: $\frac{v}{Vmax} = \frac{[S]}{(Km+[S])} \times 100 = \%\,\text{Saturation}$

$$\%\ \text{Saturation of enzyme with substrate B} = \frac{[Substrate\ B]}{Km\ (substrate\ B) + [Substrate\ B]} \times 100$$

$$V\max\ (\Delta Substrate\ A) = \frac{Vmax\ apparent\ (\Delta Substrate\ A)}{\%\ saturation\ (substrate\ B)} \times\ 100$$

The errors associated with these adjusted V_{max} values were calculated using the relationship between standard errors equation:

When $Z = B/A$

$$\left(\frac{\Delta Z}{Z}\right)^2 = \left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta B}{B}\right)^2$$

$$\text{So : } \left(\frac{\Delta Z}{Z}\right) = \sqrt{\left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta B}{B}\right)^2}$$

2.13.7.3 Substrate inhibition

Some of the enzyme activities observed in this project appeared to be under the influence of substrate inhibition. Kinetic parameters were deduced from these data through analysis using the Origin computer programme (Originlab, USA) fitting the data to the substrate inhibition equation shown below:

$$v = \frac{Vmax\ [S]}{Km + [S] + \left({}^{[S]^2}/_{Ki}\right)}$$

(where *K*i = the dissociation constant of the non-productive enzyme-substrate complex)

3 CHARACTERISATION OF THE ADHE ENZYME FROM *GEOBACILLUS THERMOGLUCOSIDASIUS*

3.1 Introduction

ADHE proteins are bi-functional enzymes that possess both an aldDH and ADH activity on a single polypeptide. Such proteins have been well studied in a variety of different organisms including:

- Gram-positive bacteria (*Leuconostoc mesenteroides (Koo et al. 2005)*, *Thermoanaerobacter ethanolicus* (Peng et al. 2008), *Streptococcus bovis* (Asanuma et al. 2004), *Clostridium acetobutylicum* (Fontaine et al. 2002)),
- Gram-negative bacteria (*E. coli* (Membrillo-Hernandez et al. 2000)),
- the anaerobic protozoan *Entamoeba histolytica* (Bruchhaus and Tannich 1994),
- the *Polytomella sp* of algae (Atteia et al. 2003) (within the mitochondrion).

At the start of this project, the *adhE* gene product from *Geobacillus thermoglucosidasius* was predicted to be the enzyme responsible for catalysing ethanol production from acetyl-CoA produced earlier in fermentative metabolism. This prediction was based upon work carried out by TMO Renewables, to look at the ability of *adhE* knock-out strains to produce ethanol under anaerobic conditions (TMO Renewables 2009 personal communication). Genes coding for several other independent ADH and aldDH proteins are annotated within the TM242 genome. From the preliminary experiments carried out by TMO Renewables, it would appear that it is the ADHE protein that drives ethanol production.

This chapter describes the work carried out to characterise the ADHE enzyme from *Geobacillus thermoglucosidasius* produced by the native organism and in recombinant expression strains of *E. coli.*

As discussed in the Introduction to this thesis, the *adhE* gene has been suggested to be under the control of an anaerobic regulation system known as REX (Sickmier et al. 2005; Wang et al. 2008). The native material chosen for assays of ADHE was from fermentatively grown cells where anoxic conditions would drive ADHE expression due to the predicted anaerobic regulation of the *adhE* gene.

The ADHE protein has been shown by several authors to form large multimeric protein complexes known as spirosomes. These are large “rod like” helical assemblies of the

ADHE monomeric unit. Kessler et al. used electron microscopy to demonstrate that the addition of ligands (NAD^+ and CoA-SH) appeared to have a dynamic effect on the overall conformation of the structures (Kessler et al. 1992). The final section of this chapter describes the work carried out to determine if such structures are present in the ADHE protein studied here.

3.2 Materials and methods

3.2.1 TM242 fermenter growth protocol

TM242 fermentation cell pellets used for this section of the project were produced in a Braun Biostat B fermenter with 2.5 litre capacity using the following protocol:

- Seed cultures were prepared using the method described in Section 2.5.2 in USM medium (30 g/L glycerol, 2% yeast extract). The fermentation was carried out in USM medium (60 g/L sucrose, 2% yeast extract).
- The sterilised fermentation vessel was inoculated with 2 L of filter sterilised medium prior to fermentation. The sterility of the system was confirmed by setting the system to 30°C, stirring at 200 rpm for 16 h prior to the fermentation run.
- The system was then set to 60°C stirring at 600 rpm, at pH 6.7 and allowed to equilibrate. The control solutions used to regulate conditions were 5 M KOH, 5 M H_3PO_4 and antifoam 204 (Sigma).
- An inoculum of 200 AU of the seed culture was added to the fermentation vessel and an OD_{600nm} measurement was taken.
- Oxygen levels were monitored using a calibrated oxygen probe. Once the level had reached 0%, the fermentation was allowed to continue for a further 2.5 h prior to ending the run.
- Samples were taken at regular time points and the OD_{600nm} measured to monitor the growth of the culture.
- Following completion of the fermentation, cells were harvested by centrifugation at 5300 x *g* for 20 min prior to snap freezing in liquid nitrogen and storing at -20°C. Typically a fermentation run would yield approximately 30 g of cell pellet.

A typical fermentation run data set and growth curve are shown in Figure 3-1.

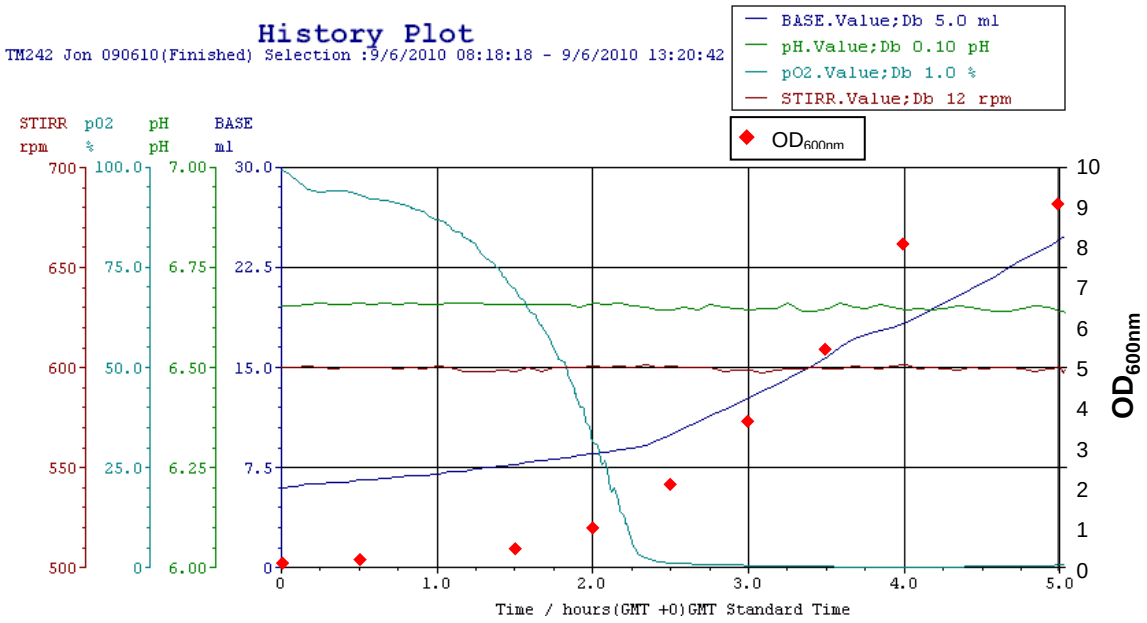


Figure 3-1: Typical fermentation run data for a TM242 culture.

3.2.2 Protein expression

Protein expression was carried out as described in Section 2.4.

When using the Arctic Express© strain, the expression culture was grown in the presence of selection antibiotics (Strep Gent Kan) at 30°C to the required OD_{600nm}. The temperature of the culture was then lowered to 12°C prior to the addition of IPTG. The protein expression was allowed to progress for 24 h prior to harvesting the bacterial cell cultures.

It is hypothesised that the slower rate of protein expression at the lower temperature used for this strain, coupled with the presence of molecular chaperonin complexes, helps to improve the ratio of soluble to insoluble protein produced.

3.2.3 Mass spectrometry analysis of SDS-PAGE gel excised protein bands

The selected SDS-PAGE gel protein bands were excised with clean razor blades and placed in individual Eppendorf tubes with 10 µl of destain solution. The MS analysis was carried out at the BSRC Mass Spectrometry and Proteomics Facility (St Andrews University). This work followed the standard protocols outlined below (Botting C.H 2012):

In-gel digestion

The gel band was cut into 1 mm cubes. These were then subjected to in-gel digestion using a ProGest Investigator in-gel digestion robot (Genomic Solutions, Ann Arbor, MI) using standard protocols. Briefly, the gel cubes were destained by washing with acetonitrile and subjected to reduction and alkylation before digestion with trypsin at 37°C. The peptides were extracted with 10% formic acid (Shevchenko et al. 1996).

MALDI TOF/TOF analysis

The digest solution (0.5 mL) was applied to the MALDI target along with alphacyano-4-hydroxycinnamic acid matrix (0.5 mL, 10 mg/mL in 50:50 acetonitrile:0.1% TFA) and allowed to dry. MALDI MS was acquired using a 4800 MALDI TOF/TOF Analyser (Applied Biosystems, Foster City, CA) equipped with an Nd:YAG 355 nm laser and calibrated using a mixture of peptides. The most intense peptides (up to 15) were selected for MS-MS analysis and the combined MS and MS-MS data were analysed, using GPS Explorer (Applied Biosystems) to interface with the Mascot 2.1 search engine (Matrix Science, London, UK), against the UniProt (Swiss-Prot and TREMBL combined) database (April 2009). No species restriction was applied. The data were searched with tolerances of 100 ppm for the precursor ions and 0.5 Da for the fragment ions, trypsin as the cleavage enzyme, assuming up to one missed cleavage, carbamidomethyl modification of cysteines as a fixed modification and methionine oxidation selected as a variable modification.

3.2.4 Dynamic light scattering

Dynamic light scattering analysis was carried out using a Nano-S Zetasizer, (Malvern, UK). 50 µl of sample was used for analysis in a low-volume quartz cuvette. Readings were taken over an 80-s period (carried out at 25°C). The buffers used were assumed to have the same viscosity as water. The analysis software on the instrument was used to provide an estimate of the protein M_r. This estimate is based on the protein being globular in nature so the results obtained should be treated with caution.

3.2.5 NanoSight analysis

Samples of purified protein were subjected to analysis using a NanoSight LM10 instrument (NanoSight, UK).

This technique monitors the scattering of laser light by particles within a suspension. By monitoring this scattering an estimate of particle size based on Brownian motion measurements (using the equation shown below) is made. This technique employs

single-particle tracking rather than providing an average of all the particles present in a sample to give an accurate representation of the particle species present. A CCD camera is used to collect images of the scattered light and these images are analysed frame-by-frame to determine the average speed of each monitored particle. This technique has been shown to be comparable to DLS in detection of nanoparticles (Filipe et al. 2010). The equation used to determine the hydrodynamic radius is based upon the Stokes-Einstein equation:

$$\overline{(x,y)}^2 = \frac{2k_B T}{3r_h \pi \eta}$$

Where $\overline{(x,y)}^2$ is the average speed of the particle squared, *T* is temperature, k_B is the Boltzmann constant, η is the viscosity of the medium and r_h is the hydrodynamic radius.

Figure 3-2 provides an overview of the technique. The ideal detectable particle number was 10^7-10^9/ml, so samples were diluted into this range with the appropriate storage buffer where required.

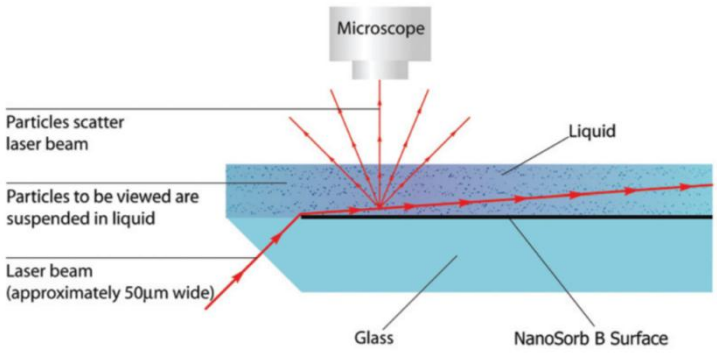


Figure 3-2: Overview of the NanoSight technology (image taken from (Malloy and Carr 2008)).

Protein samples were injected into the sample chamber and subjected to the 635 nm laser light; images were collected at a rate of 30 frames per second. The data were then analysed using the analytical software which generated the particle size distribution plots.

3.2.6 Transmission electron microscopy (TEM)

Partially-purified protein samples were adsorbed onto UV-exposed, pioloform-coated 300-mesh copper grids which had been previously coated with evaporated carbon (2 thin strands). Samples were applied to the grid and left to settle for 1 min prior to removal of the excess liquid using filter paper. Negative-stain solution was added onto the grid and was partially removed with filter paper, prior to air drying the sample. The

samples were then visualised using a JEOL (Japan) JEM1200EXII transmission electron microscope (running at 100 kV). 2% Uranyl-Acetate and 2% phosphotungstic acid (PTA) were both tried as negative-staining agents for the samples. These experiments were carried out with Mrs Ursula Potter (University of Bath Microscopy and Analysis Suite).

3.3 Native ADHE

3.3.1 Results

The following section of this chapter will focus on the study of the ADHE protein when it was produced by *Geobacillus* strains.

3.3.2 Assay development

A standard buffer was used for the enzyme assays in an effort to standardise the kinetic parameters obtained for the activities described in this project. The standard buffer used was 50 mM citric acid pH 6.0, 0.1 mM zinc acetate. Assays at 60°C were carried out in TM242 cell extracts to determine the optimum conditions for both the aldDH and ADH domains of the protein, with respect to pH, metal ions used and metal ion concentration.

Assays were carried out between pH 5.4 and 6.8 (citric acid buffer); the optimum pH for the aldDH assay was pH 6.0, whereas the optimum pH for the ADH activity was pH 6.4.

Preliminary experiments suggested divalent metal ions had a positive effect on the activity of the ADH domain. The aldDH activity was unaffected by the presence/absence of divalent metal ions. Assays were undertaken with different divalent metal ions and the effect on enzyme activity measured. The effect of the concentration of zinc present in the assays was also measured.

Metal ion (0.1 mM)	% ADH activity (of activity with no metal ions)
None	100
Zn^{2+}	144
Cu^{2+}	314
Fe^{2+}	132
Ni^{2+}	106
Co^{2+}	140
Mn^{2+}	118
Mg^{2+}	182

Table 3-1: Relative effect of metal ions in the assay buffer on ADH enzyme activity.

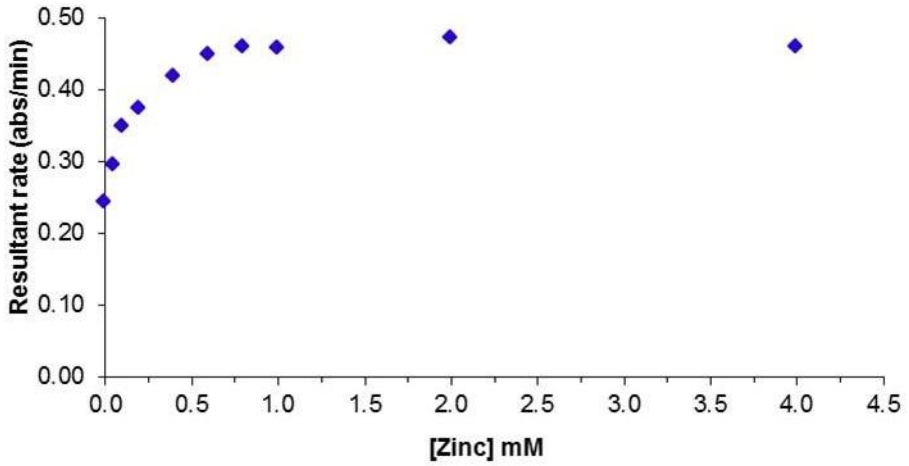


Figure 3-3: Assays of ADH activity relative to zinc acetate concentration in citric acid buffer.

No aldDH activity was detected when assayed with NADPH, suggesting this domain of the enzyme is specific for NADH. ADH activity was detected at approximately 30% of the activity observed with the same concentration of NADH, suggesting this domain is significantly more specific for NADH.

3.3.3 Optimum temperature of native ADHE activity

The optimum temperature of the ADHE enzyme activities in cell extracts was determined by assaying the enzymes using the standard assays at varying temperatures, as described in Section 2.13.5.

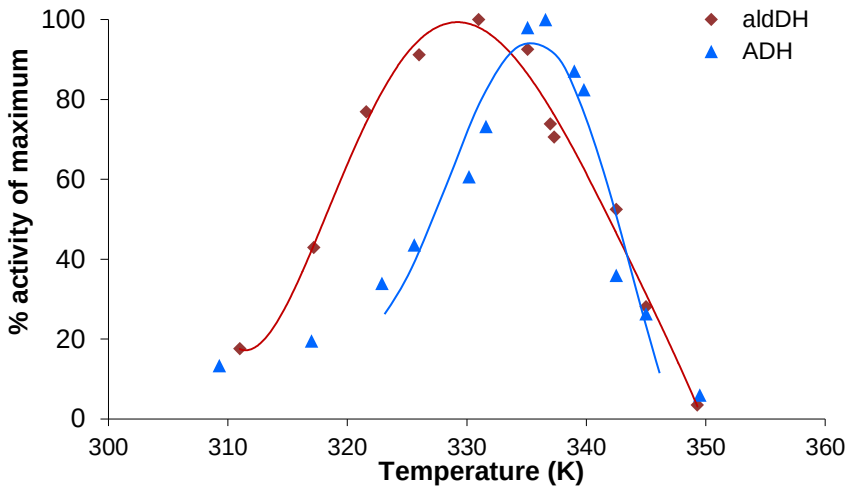


Figure 3-4: Plot of % activity of maximum against temperature (K) for both ADHE activities in native TM242 cell extracts.

The optimum temperature for the ADH activity was approximately 64°C and the optimum for the aldDH activity was approximately 58°C.

3.3.4 Tube fermentation analysis

Tube fermentations of TM242 and TM400 (ADHE k/o) strains were carried out in USM medium (111 mM glucose, 0.5% yeast extract) as described in Section 2.5.1. Metabolite concentrations were then analysed at TMO Renewables using HPLC.

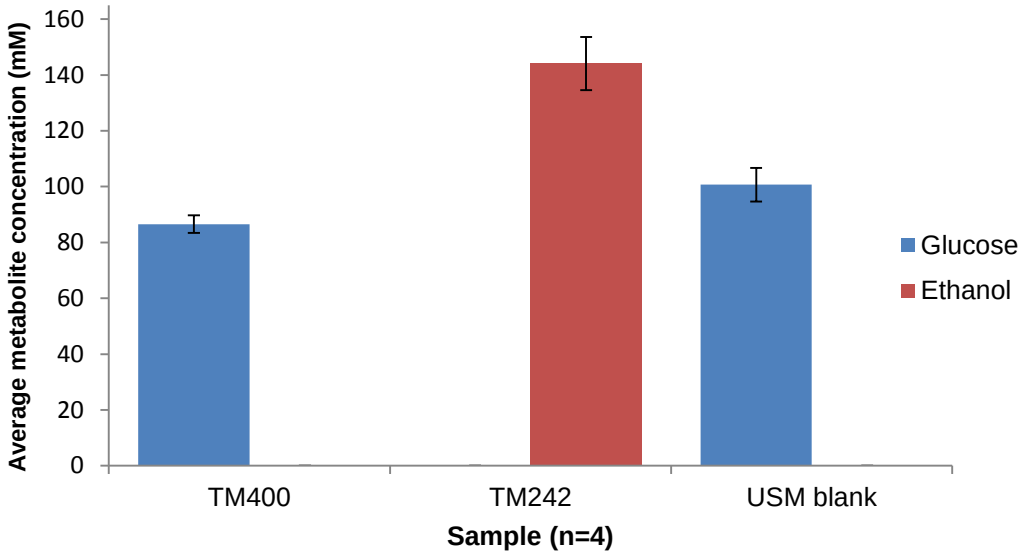


Figure 3-5: Tube fermentation analysis of TM400 and TM242 grown in USM medium. Glucose and ethanol levels are shown following 48 h growth. Results are averages of 4 replicates (error bars = standard errors).

3.3.5 SDS-PAGE analysis

The total and soluble fractions of a cell extract made from TM242 fermenter cell pellets (as described in Section 2.6) were analysed by SDS-PAGE as shown in Figure 3-6. A distinct protein band corresponding to the predicted molecular weight of ADHE was observed, and appeared to be relatively intense compared to some other protein components of the cell extract. This gives an indication that this protein may be highly expressed in this organism when grown under fermentative conditions.

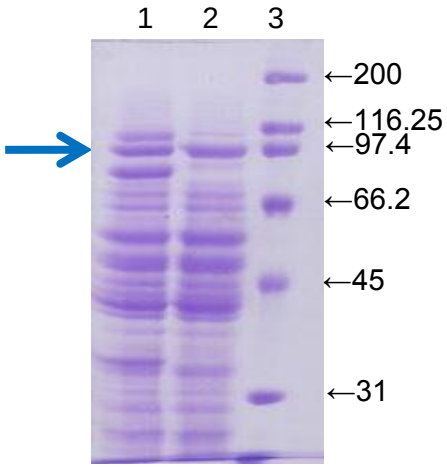


Figure 3-6: SDS-PAGE analysis of TM242 cell extract. 1 = total, 2 = soluble, 3 = markers (M_r/1000). The predicted ADHE band (M_r = 96,290) is highlighted with a blue arrow.

3.3.6 Kinetics of ADHE in cell extracts

Both ADH and aldDH enzymatic activities were detected in TM242 cell extracts. Accurate kinetic analysis was undertaken to obtain estimates of the key kinetic parameters for these two activities from the native organism. A substrate-independent NADH-oxidase background activity was detected in the cell extract. This was taken into account when determining the observed rate of reaction during these assays. Where appropriate, the data were analysed by the direct linear method (Eisenthal and Cornish-Bowden 1974) using the Enzpack computer programme (Biosoft).

3.3.6.1 AldDH

Results of the aldDH activity assays (using the aldDH NADH assay, Section 2.13.1.2) are shown in the following Figures.

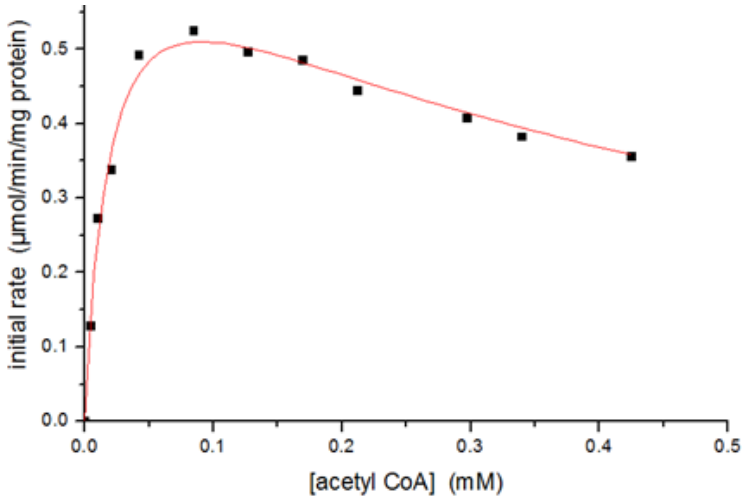


Figure 3-7: Michaelis-Menten plot of aldDH enzymic activity (U mg^{-1} of protein) against concentration of acetyl-CoA (mM) at a fixed concentration of NADH (0.22 mM) in a TM242 cell extract.

Non Michaelis-Menten kinetics were observed for the aldDH activity with respect to acetyl-CoA. These data appear to fit the trend expected for substrate inhibition. To allow kinetic parameters for NADH to be determined, a fixed concentration of acetyl-CoA, close to that where the maximum observed activity was obtained was used (0.13 mM).

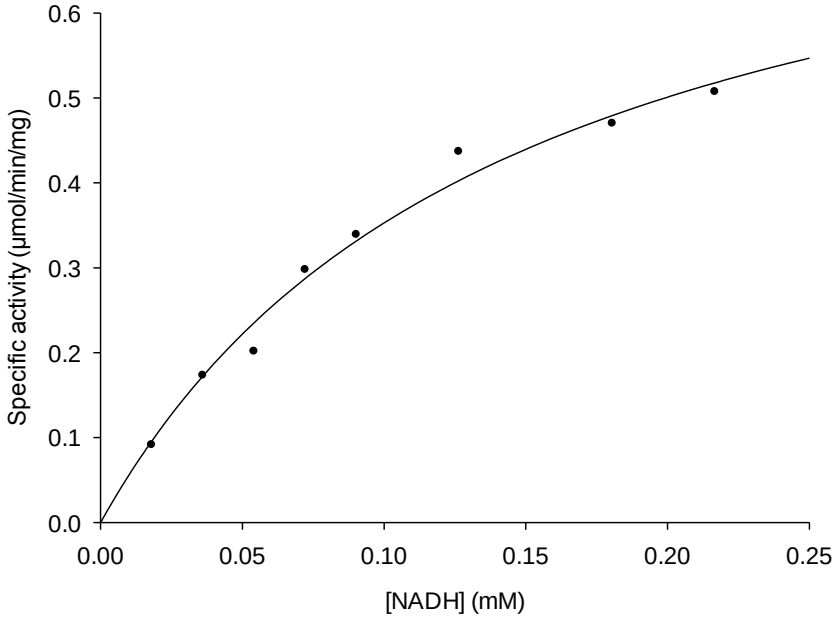


Figure 3-8: Michaelis-Menten plot of aldDH enzymic activity (U mg^{-1} of protein) against concentration of NADH (mM) at a fixed concentration of acetyl-CoA (0.13 mM) in a TM242 cell extract.

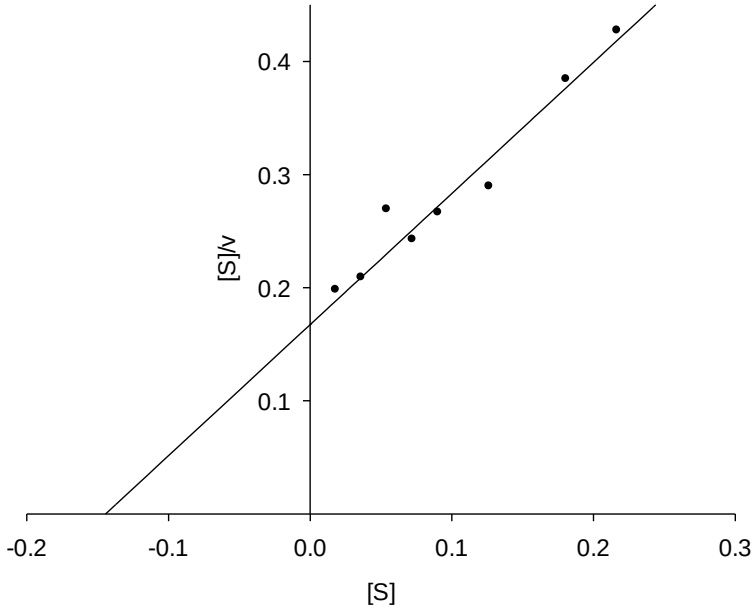


Figure 3-9: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of aldDH activity (U mg^{-1}) with respect to concentration of NADH (mM).

3.3.6.2 ADH

Results of the ADH activity assays (using the ADH assay described in Section 2.13.2) are shown in the following Figures. The results of these assays with respect to varying the concentration of acetaldehyde did not appear to obey Michaelis-Menten kinetics, and appeared to show a "biphasic" relationship between V and [S]. The presence of a second ADH activity within the cell extract is a likely explanation. It was possible to resolve the kinetic parameters for each by deducting the observed V_{max} of the lower activity from the observed activities of the higher. "ADH" corresponds to the ADH activity with the higher K_m for acetaldehyde whereas "ADH2" refers to the ADH activity with a lower K_m for acetaldehyde.

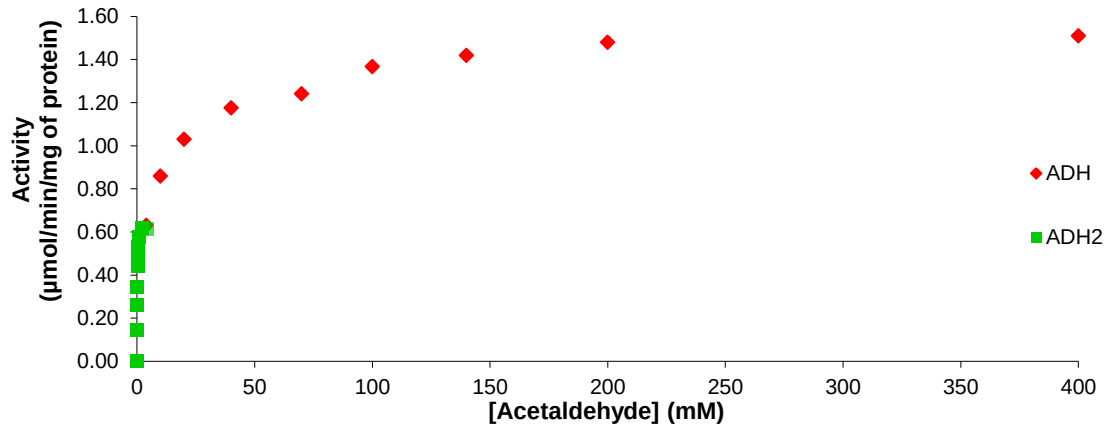


Figure 3-10: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) against concentration of acetaldehyde (mM) at a fixed concentration of NADH (0.22 mM) in a TM242 cell extract "full range".

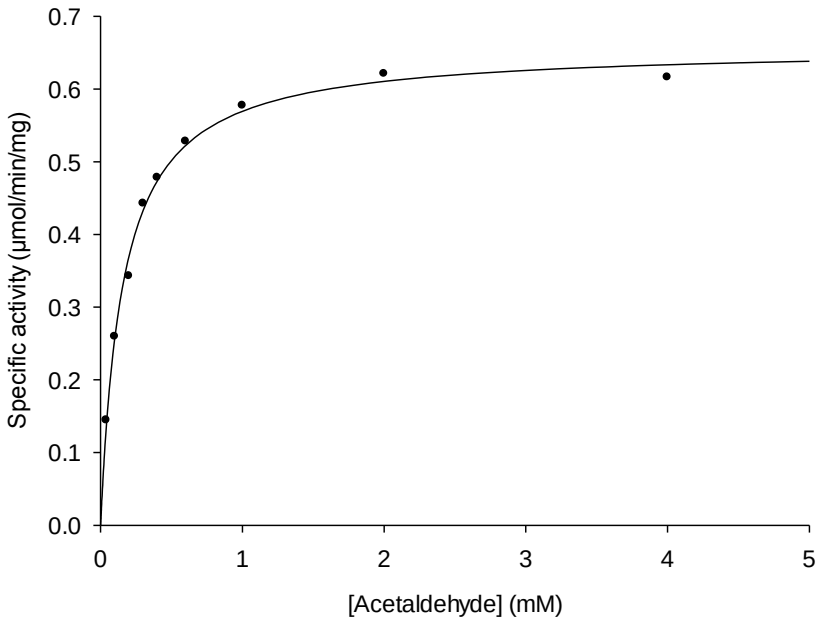


Figure 3-11: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) against concentration of acetaldehyde (mM) up to 4 mM at a fixed concentration of NADH (0.22 mM) in a TM242 cell extract.

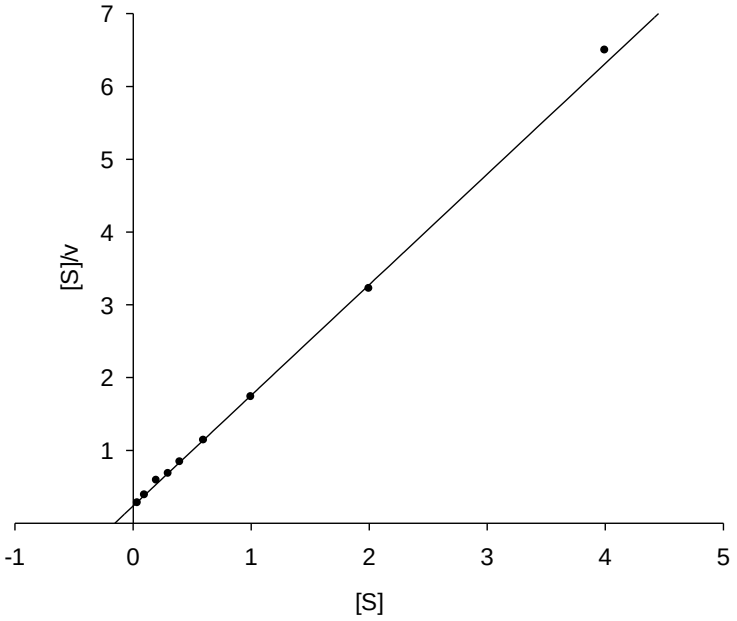


Figure 3-12: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of acetaldehyde (mM) up to 4 mM acetaldehyde).

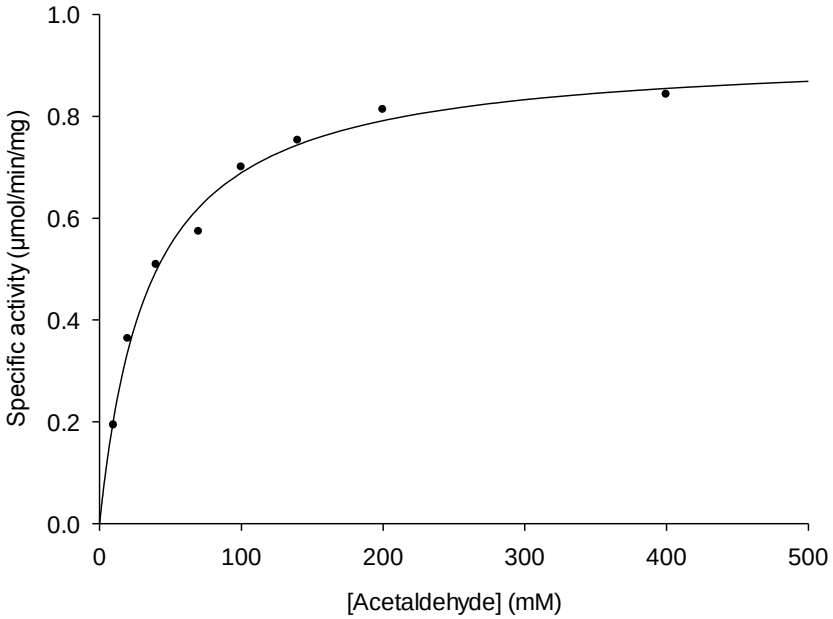


Figure 3-13: Michaelis-Menten plot of adjusted ADH enzymic activity (U mg^{-1} of protein) against concentration of acetaldehyde (mM) 10-400 mM at a fixed concentration of NADH (0.22 mM) in a TM242 cell extract.

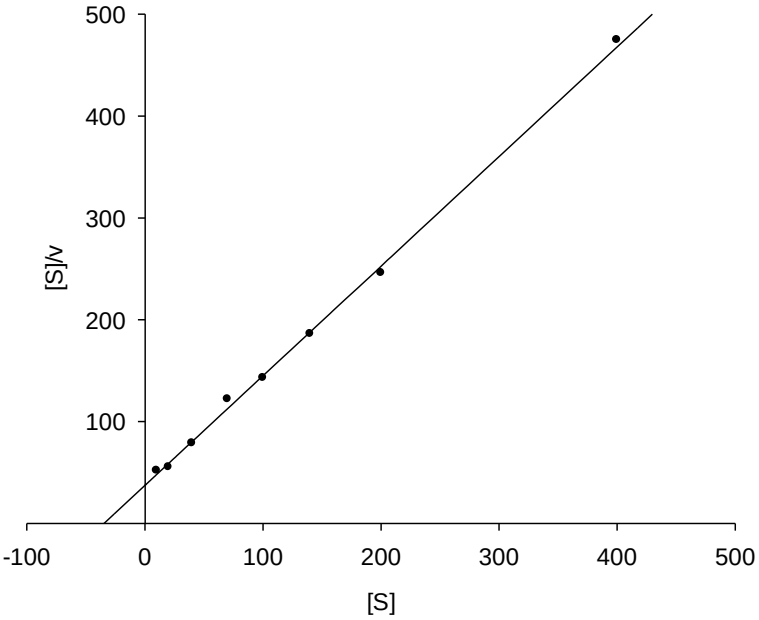


Figure 3-14: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of acetaldehyde (mM) (10-400 mM acetaldehyde).

Variable concentration of NADH ADH assays in the cell extract did not show biphasic activity, so the full range of data was used to determine the kinetic parameters.

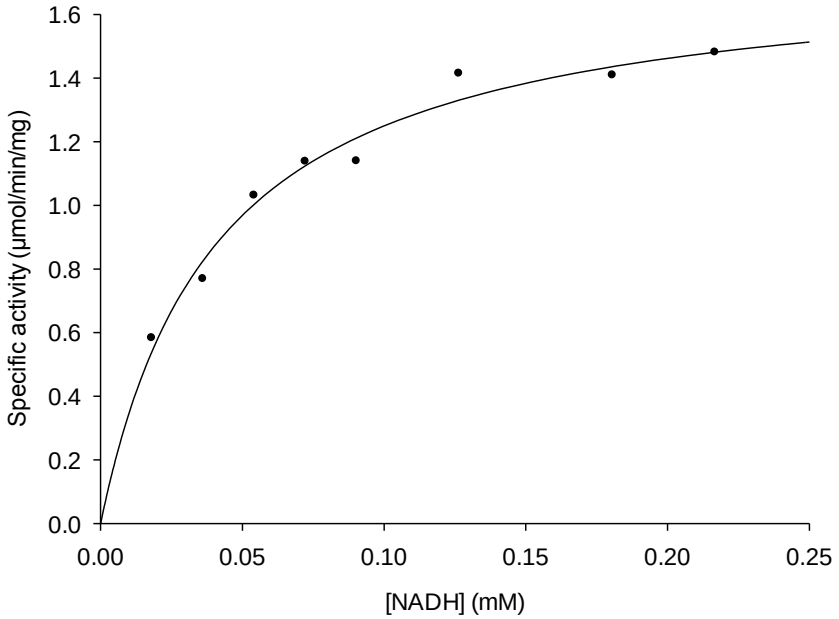


Figure 3-15: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) against concentration of NADH (mM) at a fixed concentration of acetaldehyde (200 mM) in a TM242 cell extract.

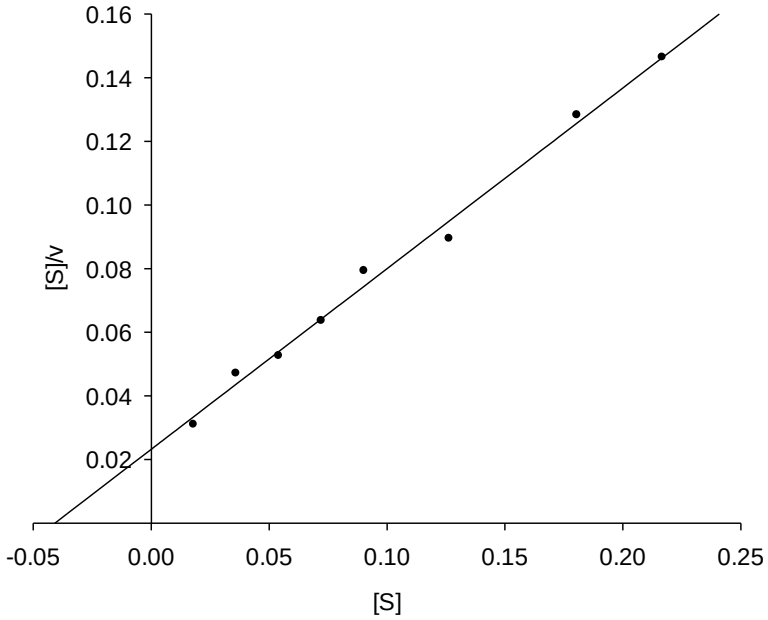


Figure 3-16: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of NADH (mM).

3.3.6.3 Summary of kinetic analyses

It was not possible to fully saturate the enzyme with the fixed substrates during these assays (as described in Section 2.13.7). Using the kinetic parameters determined, the percentage saturation was calculated for each substrate. These can then be used to adjust the observed V_{max} values to give an estimation of the true V_{max}. The errors stated take into account any adjustment made using this calculation. U is defined as µmol substrate (NADH) converted per min.

Activity	Substrate	Kinetic parameter	Parameter value		Units
ADH	Acetaldehyde	K_m	33.8	+/- 1.9	mM
		$V_{max\ (adjusted)}$	1.1	+/- 0.1	U mg^{-1}
ADH	NADH	K_m	0.038	+/- 0.003	mM
		$V_{max\ (adjusted)}$	2.0	+/- 0.1	U mg^{-1}
ADH 2	Acetaldehyde	K_m	0.160	+/- 0.002	mM
		$V_{max\ (adjusted)}$	0.78	+/- 0.06	U mg^{-1}
AldDH	Ac-CoA*	K_m	0.019	+/- 0.003	mM
		$V_{max\ (adjusted)}$	1.24	+/- 0.14	U mg^{-1}
AldDH	NADH	K_m	0.157	+/- 0.014	mM
		$V_{max\ (adjusted)}$	1.0	+/- 0.2	U mg^{-1}

Table 3-2: Summary of kinetic parameters determined in TM242 cell extracts. *Kinetic parameters for aldDH with respect to acetyl-CoA were determined by fitting the data to the substrate inhibition equation as described in Section 2.13.7.3.

As the two ADH activities could not be distinguished with respect to NADH, the V_{max} for the ADH activity shown in Table 3-2 (NADH) is based on the two ADH activities together. The predicted value from both the ADH V_{max} values together is 1.88 U mg^{-1} which is close to the observed value of 2.0 U mg^{-1}.

3.3.6.4 NAD:CoA released ratio assays

Assays of the ratio of NADH oxidised relative to CoA-SH produced were carried out on TM242 cell extracts (as described in Section 2.13.4). Background rates used in these assays took into account both the NADH oxidase (acetyl-CoA independent NADH oxidation) and any PAT (NADH-independent DTNB cleavage) activities that may be present.

The average of three separate assays gave an estimate of 1.8 NADHs oxidised per 1 CoA-SH produced. This suggests that proteins within the cell extracts were converting acetyl-CoA through to ethanol within the enzyme assays. The aldDH assays may therefore be a measure of metabolic flux to ethanol.

3.3.7 Preliminary substrate specificity experiments

Preliminary experiments were carried out to assess the substrate specificity of the ADHE activity in TM242 cell extracts. AldDH activity was observed with both butyryl-CoA and propionyl–CoA. It was not possible to determine the concentration of these substrates, so accurate kinetic parameters could not be obtained. Assuming the concentration of these acyl-CoAs were comparable to that of the acetyl-CoA prepared at the same time, the specific activities obtained with a fixed concentration of each acyl-CoA (0.054 mM) were compared.

Substrate	Specific activity (U mg^{-1})	% activity of Acetyl-CoA
Acetyl-CoA	0.57	100.0
Propionyl-CoA	0.37	64.9
Butyryl-CoA	0.34	59.6

Table 3-3: AldDH activity with different acyl-CoAs in TM242 cell extracts.

ADH activity with propionaldehyde and butyraldehyde was also compared to acetaldehyde. Due to the immiscibility with water of these compounds, 1 µl of neat compound was added directly to the cuvette. Comparable levels of enzyme activity were observed for all three aldehydes.

Substrate	Resultant rate (abs/min)	% activity of Acetaldehyde
Acetaldehyde	0.26	100.0
Propionaldehyde	0.24	92.3
Butyraldehyde	0.22	84.6

Table 3-4: ADH activity with different aldehydes in TM242 cell extracts.

3.3.8 Purification of native ADHE

Attempts were made to isolate the ADHE protein from TM242 using anion exchange chromatography and gel filtration. The objective of these experiments was to resolve the ADHE protein from other proteins in the cell extract but also to estimate the molecular weight of the protein in solution.

The isoelectric point (pI) of the native ADHE protein was calculated to be 7.11 using the Protparam tool (ExPASy).

3.3.8.1 Initial purification attempts

A wide variety of different purification techniques were utilised in an effort to resolve the native ADHE protein from cell extracts. Most effort was focussed on using separation by charge (anion exchange) followed by size exclusion (gel filtration) under a variety of different conditions. These purifications were carried out on an ÄKTA Explorer FPLC system (GE Healthcare).

During these purifications the aldDH activity of the protein appeared to be unstable relative to the ADH activity. To confirm whether this loss of activity was due to the removal of a second aldDH activity from within the cell extract, every fraction from an anion exchange run of a TM242 cell extract was assayed for aldDH activity. No other peak of aldDH activity was observed other than that co-eluting with ADH activity corresponding to ADHE. The presence of reducing agents (DTT and reduced glutathione (GSH)) appeared to improve the stability of the aldDH activity of the protein through purifications. These results suggest a potential oxygen sensitivity of the aldDH domain. The addition of EDTA appeared to have a negative effect on the activity of the ADH domain of the protein.

Other purification strategies included salt precipitation and dye-binding methods. Ammonium sulphate precipitation was carried out between 40 & 70% saturation; however this technique was not specific enough for ADHE. Dye-binding purification

was also attempted using a Reactive Red agarose column eluting with KCl; again, many other proteins co-eluted with ADHE using this technique.

3.3.8.2 Optimised ADHE purification

A TM242 soluble cell extract (~6 ml) (prepared as described in Section 2.6 replacing 50 mM EPPS pH 8.0, with anion exchange buffer A) was loaded onto two pre-equilibrated GE HiTrap 5 ml Q-Sepharose HP columns run in series at a flow rate of 1 ml/min. Buffer A consisted of 50 mM EPPS pH 8.7, 0.1 mM zinc acetate and 5 mM reduced glutathione. Proteins were eluted from the columns using a 0-1 M NaCl gradient (in buffer A) over a 60-min period. Fractions were assayed to identify the peak of ADHE activity prior to pooling the highest activity fractions ready for the following purification step.

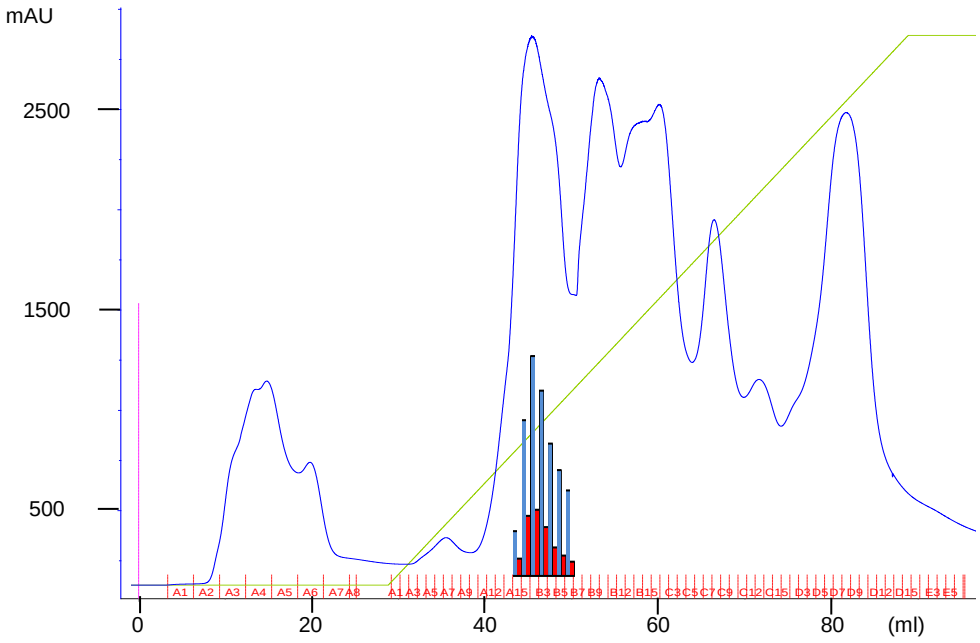


Figure 3-17: Chromatograph of a TM242 cell extract anion exchange purification. Blue line = A_{280nm} (mAU), Green line = salt gradient of 0-1 M. Activity measurements (abs/min) for both ADH and aldDH are overlaid onto corresponding fractions (blue = ADH, red = aldDH). Fractions A15-B6 were pooled, assayed and concentrated prior to gel filtration.

5.5 ml of the pooled anion-exchange sample (fractions A15-B6) was concentrated down to 0.75 ml using a Vivaspin 5K MWCO centrifugal filter device (Sartorius). 0.6 ml of this concentrated pool was loaded onto a pre-equilibrated GE Superdex 200 10/300 GL column run at a flow rate of 0.3 ml/min. The buffer consisted of 50 mM EPPS pH 8.0, 0.1 mM zinc acetate, 5 mM reduced glutathione, and 10 % (v/v) glycerol.

Fractions were assayed to identify the peak of ADHE activity prior to pooling the highest activity fractions.

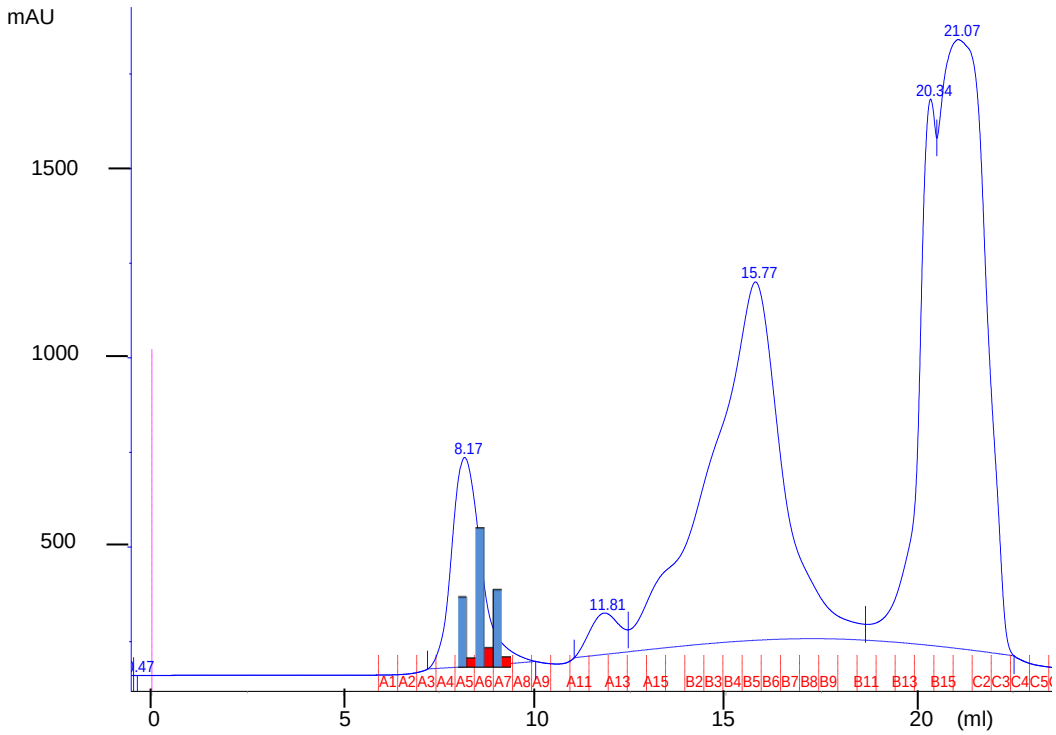


Figure 3-18: Chromatograph of a TM242 anion exchange sample gel filtration purification. Blue line = A_{280nm} (mAU). Activity measurements (abs/min) for both ADH and aldDH are overlaid onto corresponding fractions (blue = ADH, red = aldDH). Peak activity fractions (A5-A7) were pooled and assayed.

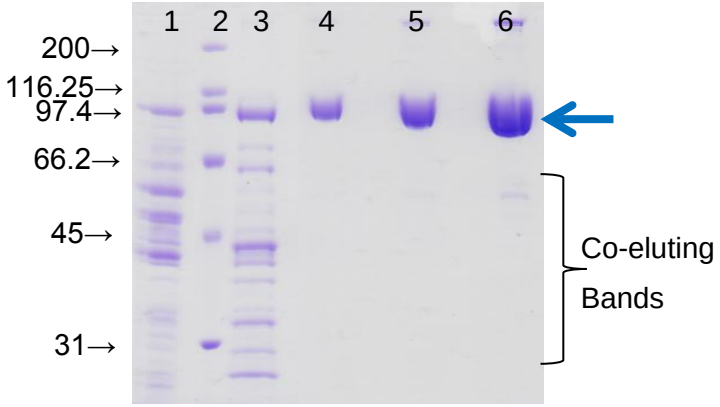


Figure 3-19: SDS-PAGE analysis of ADHE purification from TM242 cell extracts. 1 = soluble, 2 = markers (M_r/1000), 3 = pooled anion exchange sample, 4-6 = pooled gel filtration sample run at different protein concentrations. Predicted ADHE band is highlighted with a blue arrow. Several other proteins appeared to co-elute with ADHE from the gel filtration column; these are not very clear on the SDS PAGE gel image but are indicated.

Assay	Sample	Activity (µmol/min/ml)	Total activity (µmol/min)*	Protein concentration (mg/ml)	Specific Activity (U mg^{-1})	% recovery
ADH	Soluble	24.4	146.6	12.7	1.9	100.0
	Anion Exchange	13.9	83.8	1.9	7.4	57
	Gel Filtration	23.7	56.4	0.5	49.3	38
Assay	Sample	Activity (µmol/min/ml)	Total activity (µmol/min)*	Protein concentration (mg)	Specific Activity (U mg^{-1})	% recovery
AldDH	Soluble	7.1	42.4	12.7	0.6	100.0
	Anion Exchange	3.1	18.7	1.9	1.6	44
	Gel Filtration	2.7	6.5	0.5	5.7	15

Table 3-5: Protein purification table for ADHE purification from native cell extract, *gel filtration total activity has been adjusted to reflect theoretical activity had previous step's entire sample been loaded onto column.

3.3.9 Estimation of the M_r of the ADHE protein

The peak of ADHE activity eluted from the gel filtration column in the void volume. This provided an estimate of M_r of greater than 1.3×10^6 which would correspond to a protein complex of larger than 13 ADHE polypeptide units (9.6×10^4 per ADHE).

3.3.10 Kinetics of partially-purified ADHE

To determine if the second ADH activity had been resolved through this purification strategy, the partially-purified ADHE protein from TM242 cell extract was subjected to kinetic analysis. The biphasic activity was still present but the relative levels of the two activities had changed significantly. The kinetic parameters for each activity could once again be resolved by deducting the observed V_{max} of the lower activity from the observed activities of the higher.

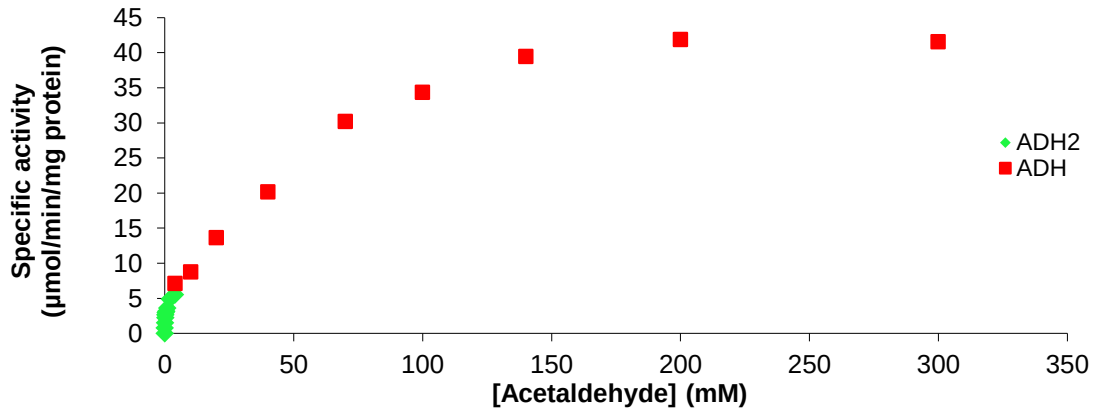


Figure 3-20: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) against concentration of acetaldehyde (mM) at a fixed concentration of NADH (0.21 mM) in a partially-purified TM242 cell extract.

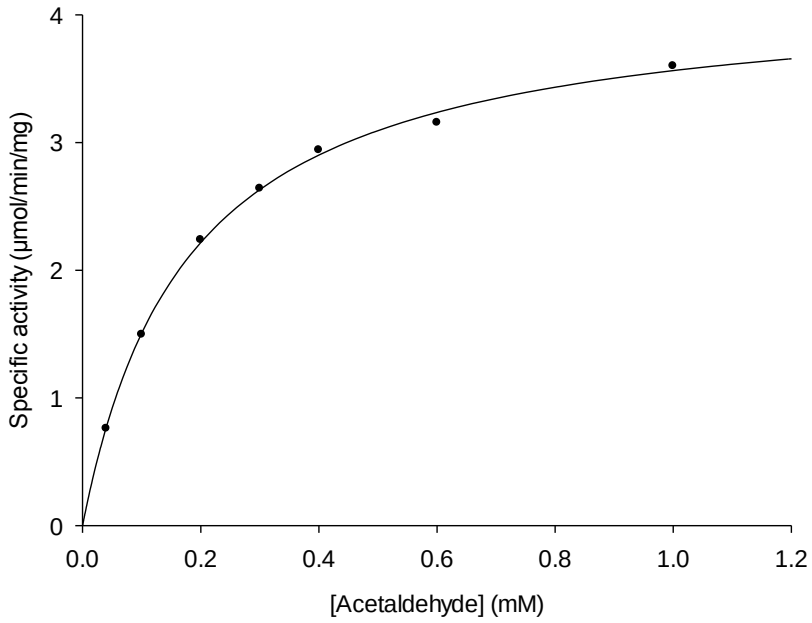


Figure 3-21: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) against concentration of acetaldehyde (mM) up to 1 mM at a fixed concentration of NADH (0.21 mM) in a partially-purified TM242 cell extract.

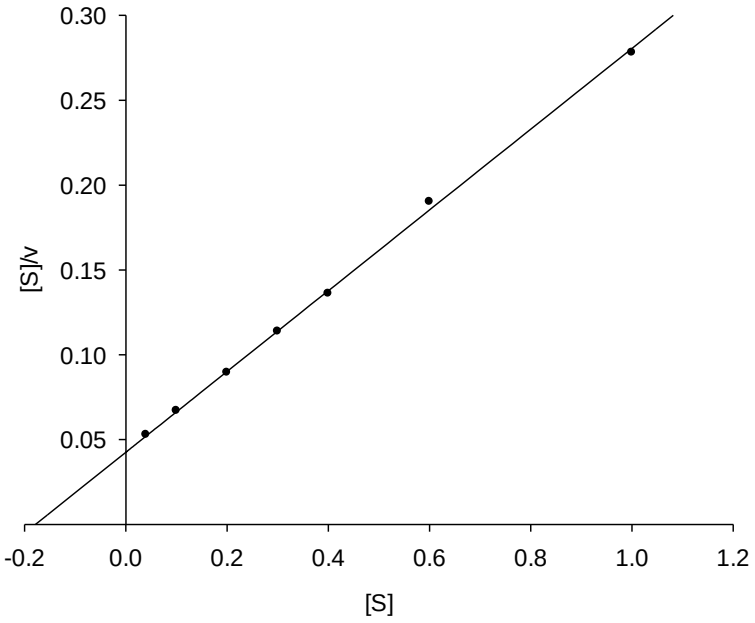


Figure 3-22: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of acetaldehyde (mM) (up to 1 mM acetaldehyde).

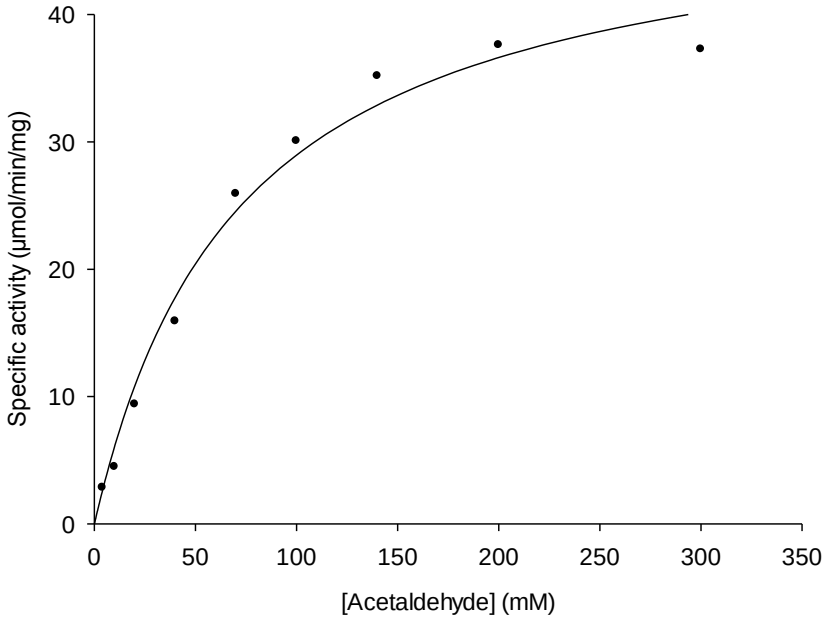


Figure 3-23: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) against concentration of acetaldehyde (mM) 4-300 mM at a fixed concentration of NADH (0.21 mM) in a partially-purified TM242 cell extract.

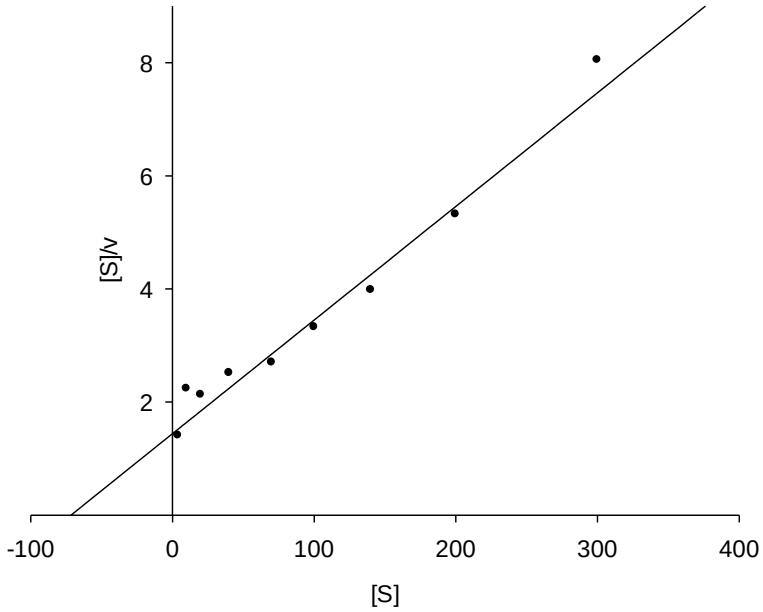


Figure 3-24: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of acetaldehyde (mM) (4-300 mM acetaldehyde).

Activity (Acetaldehyde)	**K_m (mM)**	**Standard Error**	**V_{max} (U mg^{-1})**	**Standard Error**	**Relative % activity [(V_{max} ADH2)/(V_{max} ADH)]x100**
"ADH" CE	33.8	+/- 1.9	**0.93**	+/- 0.01	71.9%
"ADH 2" CE	0.155	+/- 0.002	**0.67**	+/- 0.002	
"ADH" GF	79.5	+/- 6.6	**51.5**	+/- 2.1	8.3%
"ADH 2" GF	0.184	+/- 0.002	**4.26**	+/- 0.03	

Table 3-6: Comparison of kinetic parameters determined for ADH activities in TM242 cell extract and in partially-purified sample. (V_{max} values not adjusted for saturation). CE = Cell Extract GF = sample following gel filtration. Relative % activity corresponds to the "ADH2" activity relative to the "ADH" activity.

The kinetics of the aldDH activity for the partially-purified enzyme with respect to acetyl-CoA concentration were also determined:

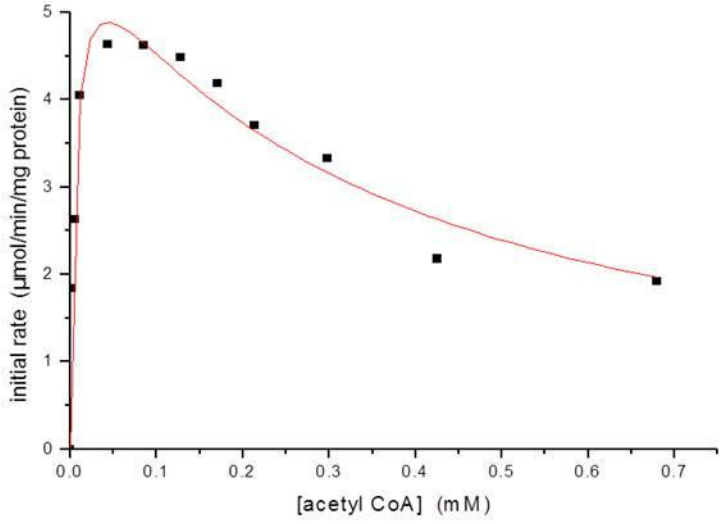


Figure 3-25: Michaelis-Menten plot of aldDH enzymic activity (U mg^{-1} of protein) against concentration of acetyl-CoA (mM) at a fixed concentration of NADH (0.22 mM) in a partially-purified TM242 cell extract.

Activity	Substrate	Kinetic parameter	Parameter value		Units
AldDH	Ac-CoA*	K_m	0.006	+/- (0.001)	mM
		V_{max}	6.22	+/- (0.38)	U mg^{-1}

Table 3-7: Kinetic parameters determined for partially-purified ADHE aldDH domain. *Kinetic parameters for aldDH with respect to acetyl-CoA were determined by fitting the data to the substrate inhibition equation as described in Section 2.13.7.3.

3.3.11 TM400 enzyme assays

TM400 was grown using the overnight shake-flask method (Section 2.5.2) to produce aerobically grown cell pellets for assays. A soluble cell extract was produced by the standard method. An ADH activity independent of aldDH activity was detected in the cell extract and the activity was shown to be proportional to enzyme concentration. The specific activity for this ADH was determined to be 0.046 U mg^{-1}, however, due to the low level of this activity and a relatively high NADH oxidase background, accurate kinetic parameters were not determined.

3.3.12 Discussion

The standard assay buffer used for the assays in this project was based on a standardised ADH assay buffer. Zinc is commonly associated with alcohol dehydrogenase proteins (BRENDA enzyme database 2012) as both a structural and a catalytic cofactor, and was initially found to cause a higher activation of the ADH activity of the protein than the other metal ions originally tested. Towards the end of

the experimental phase of this thesis it was discovered that copper and magnesium appeared to give a higher activation than zinc.

During the initial stages of the project, the buffer used for the assays was changed to citric acid buffer due to interference from the activity of the PAT protein (phosphate acetyl-transferase) in phosphate buffers when using TM242 cell extracts. Reassessment of metal ion concentration (again towards the end of the experimental phase of the project) suggested that in citric acid higher concentrations of zinc should have been used to achieve optimal activity as the buffer is known to chelate metals ions (AppliChem 2008). As all the assays were carried out in this citric acid buffer, the results obtained are still comparable but it should be noted that these values may not be optimal for ADH.

The work carried out during this section of the chapter has successfully confirmed the important role of the ADHE protein within the TM242 strain.

Tube fermentation experiments confirmed the findings of TMO Renewables (TMO Renewables 2009 personal communication) showing that when the organism does not have the ADHE protein it not only loses the ability to produce ethanol, but also the ability to metabolise glucose efficiently under anaerobic conditions. SDS-PAGE analysis of fermenter-grown cell extracts indicated that the ADHE protein may be highly expressed under fermentative growth conditions.

Kinetic analysis of the soluble cell extract of TM242 showed that metabolic flux to ethanol was occurring during the reaction in the cuvette. This result suggests that the estimates of aldDH maximal activity may be an overestimate as NADH-linked assays were used during these experiments and it is likely that 2 NADHs were converted per CoA-SH released. That is, the product of aldDH, acetaldehyde, appears to be converted to ethanol by the ADH activity of ADHE, resulting in a further NADH molecule being oxidised to NAD^+.

The aldDH activity of ADHE appears to be subject to substrate inhibition, although the relatively high fitting errors observed for the kinetic parameters suggest other factors may also contribute to the unusual kinetics with respect to acetyl-CoA. Due to the influence of this suspected inhibition the kinetic parameters with respect to acetyl-CoA were estimated using the Origin computer programme fitting the data to the substrate inhibition equation as discussed in Section 2.13.7.3. The errors associated with this

fitting are relatively high so the estimation of kinetic parameters for this domain of the protein is also subject to these errors. Different values of K_m were determined for the aldDH in cell extract and in the partially-purified sample. The data are in a similar region of concentration (0.019 ±0.003 mM and 0.006 ±0.001 mM). These data show that the K_m for acetyl-CoA is relatively low compared to that observed for the ADH protein for acetaldehyde.

The loss of aldDH activity observed during initial purification attempts did not appear to be due to the removal of other aldDH activities from the cell extract. This was demonstrated by assaying all the fractions from an anion-exchange purification. The presence of reducing agents was shown to reduce the loss of aldDH activity through purification. A catalytic thiol residue in the active site of this domain may help explain the apparent oxygen sensitivity of this protein. Cys_{257} is highly conserved between the TM242 ADHE and several other ADHE proteins discussed in the literature as indicated in the Introduction to this thesis (Chen et al. 2004; Espinosa 2001). Oxidation of this catalytic thiol residue could be the cause of the significant loss of aldDH activity observed.

The ADH activity observed in the cell extracts appears to show a “biphasic” activity, i.e. at least two ADH activities appeared to be present. The two activities can be differentiated due to their contrasting kinetic parameters, “ADH2” with a relatively low K_m for acetaldehyde (0.155 mM) and “ADH” with a relatively high K_m (33.8 mM). A similar biphasic activity was observed in other assay buffers between pH 6-7, suggesting that it is not the result of an anomaly caused by the assay conditions.

A significant decrease in the “ADH2” activity relative to the suspected ADH activity of ADHE, was observed following the partial purification of the ADHE protein. This result supports the hypothesis that it is the presence of a second ADH activity within the cell extract causing the non Michaelis-Menten kinetics observed. Several other protein bands appeared to co-elute with the ADHE protein following purification, and it is likely that one of these bands may correspond to an independent ADH protein. Such a protein may interact with ADHE causing the two proteins to co-elute following anion-exchange and gel-filtration purifications. The difference in K_m values observed between the purified and non-purified ADH activities (79.5 mM compared to 33.8 mM) could reflect the influence of the “ADH2” activity on the deduced ADHE activity in cell extracts. As discussed in the introduction to this chapter, several other ADH-encoding

genes are annotated within the TM242 genome that may be responsible for this second activity.

TM400 cell extracts produced were assayed for ADH activity, in an effort to confirm whether the biphasic activity observed for ADH was indeed due to the presence of a second ADH protein. A low level of ADH activity was observed in the aerobically grown TM400 strain (0.046 U mg^{-1}); however, this activity was too low to determine accurate kinetic parameters. Without growing this strain under the fermentative conditions used for TM242, it is difficult to say whether this ADH activity is indeed the observed "ADH2" activity from the cell extract. The ADHE knock-out in TM400 means this strain cannot grow anaerobically, and thus cannot be used as a true control for comparison with fermentatively grown TM242. It is possible that the expression of both ADHE and this second "ADH2" are controlled by anaerobic promoters such as the REX system discussed in the Introduction.

The elution of the ADHE protein in the void volume of the gel filtration column suggests that it may form a very large protein complex within the cell extract or that the protein forms large aggregates during purification. These hypotheses will be discussed in a later section of this chapter.

The following section of this chapter will focus on the use of recombinant protein expression to study ADHE when it was not produced by *Geobacillus* strains.

3.4 Recombinant ADHE

3.4.1 Results

The plasmid pTMO259 (pET28a-*adhE*) was supplied by TMO Renewables at the start of this project. The *adhE* gene was ligated into the plasmid between the NheI/XhoI sites of the pET28a vector. This introduced an N-terminal histidine tag to the ADHE protein to facilitate purification.

3.4.2 Gene cloning and protein expression trials

The pTMO259 plasmid was transformed into competent protein expression strains and transformants were screened with the required selection antibiotics.

Protein expression trials of the ADHE protein were carried out in *E. coli* expression strains BL21 (DE3), Rosetta© (DE3) and Arctic Express© (DE3). Various attempts were made to improve the yield of soluble protein obtained, including the use of auto-

induction medium (Overnight Express, Novagen), varying the concentration of and time of exposure to IPTG and using higher temperature expression according to the work of Koma et al. (2006). The Arctic Express© (ArcX) expression strain, which possesses two psychrophilic chaperonin proteins for low temperature protein expression, was selected for use in this project as this strain appeared to significantly increase the ratio of soluble to insoluble protein.

Cells of this strain were grown according to the described method in Section 3.2.2 (lowering the temperature prior to induction from 30°C to 12°C).

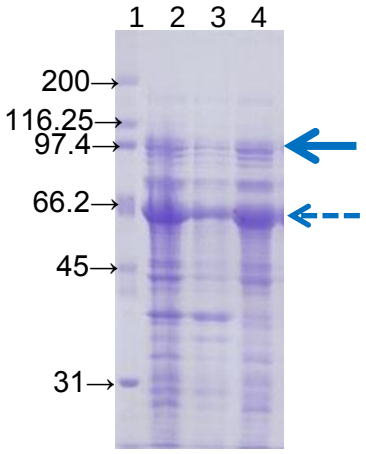


Figure 3-26: SDS-PAGE analysis of ArcX ADHE protein expression trial. 1 = markers (M_r/1000), 2 = total, 3 = insoluble, and 4 = soluble. Predicted His-tagged-ADHE band (M_r = 98,456) is highlighted with a blue arrow (samples were heavily diluted due to the presence of the overexpressed Cpn60 chaperone (indicated with dashed arrow).

3.4.3 Protein purification

Cell extracts of ArcX ADHE were prepared as described in Section 2.6. The protein expressed from pTMO259 has an N-terminal histidine tag which allowed purification using a metal-affinity chromatography column charged with Ni^{2+} (according to the method outlined in Section 2.9).

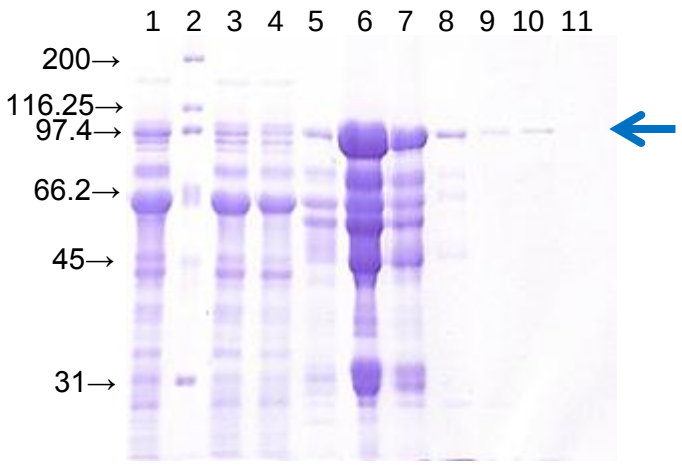


Figure 3-27: SDS-PAGE analysis of ArcX ADHE protein metal affinity chromatography samples. 1 = soluble, 2 = markers (M_r/1000), 3 = flow through, 4 = 0%, 5 = 5%, 6-10 = 1 ml 40% washes and 11 = 100%. Predicted His-tagged-ADHE band (M_r = 98,456) is highlighted with a blue arrow) (% values = %HIS-ELUTE buffer diluted in HIS-BIND buffer).

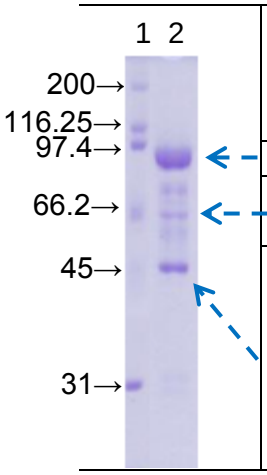
The protein product (M_r 98,456) bound to the column and was eluted with a 40% solution of HIS-ELUTE. Several smaller co-contaminant proteins appeared to be eluted with the protein of interest. The presence of imidazole used during the protein purification appeared to interfere with the assays of ADHE activity, so the 40% elution samples were pooled and dialysed into 60 mM tetrasodium pyrophosphate, pH 6.0, 0.1 mM zinc acetate ready for assays.

Following metal-affinity purification, the use of dye-binding purification was also attempted using a Cibacron blue agarose column. Attempts were made to elute the protein with various concentrations of NAD^+ and NaCl. The ADHE protein bound too strongly to the column and could not be eluted in an active form.

Gel filtration of the His-purified material was also carried out in an attempt to resolve the co-eluting bands observed, but several of the co-purified proteins remained following this purification. The recombinant ADHE appeared to elute in the void volume of the gel filtration column as was observed with the native ADHE protein.

3.4.4 Mass spectrometry (MS) of co-eluting bands

The pooled 40% HIS-ELUTE fraction was diluted and run on an SDS-PAGE gel prior to the key bands observed being excised for Matrix assisted laser desorption/ionisation-time of flight/time of flight (MALDI-TOF/TOF) mass spectrometry analysis as described in Section 3.2.3.



Sample number	Predicted protein	Mascot score	% Sequence Coverage	Number of unique peptide peaks
1	ADHE	1750	62	15
3	Cpn60	876	57	11
	ADHE	145	21	1
2	ADHE (ADH domain)	1330	50	12

Figure 3-28: SDS-PAGE analysis of purified ArcX ADHE protein and mass spectrometry analysis. 1 = markers (M_r/1000), 2 = sample with bands analysed indicated. Overloading of the gel to maximise the concentration of co-eluting bands appeared to cause the ADHE protein to run further than expected on this SDS-PAGE gel.

MS analysis revealed that sample 1 was the expected His-tagged ADHE recombinant protein. Sample 2 also corresponded to the ADHE protein expected. However, reviewing the ion scores for this sample, only 1 of the 12 unique peptide peaks were observed in the N-terminal domain of the protein; the other 11 were all present in the ADH domain of the protein. Sample 3 consisted of a chaperonin from *Oleispira antarctica* as well as the ADHE protein. The sequence coverage was poor for the ADHE protein suggesting only a small fragment of the protein was present in the sample.

3.4.5 Kinetics of partially-purified recombinant ADHE

The dialysed pooled protein sample of His-purified recombinant ADHE material was used for kinetic characterisation. ADH and aldDH enzymatic activities were detected in the purified protein sample. Accurate kinetic analysis was undertaken to obtain estimates of the key kinetic parameters for these two activities. Where appropriate, the data were analysed using the direct linear method using the Enzpack computer programme (Biosoft).

3.4.5.1 AldDH

Following the addition of enzyme to start the reaction, a lag in activity was observed for the recombinant aldDH activity. During these experiments the aldDH assay was modified to measure the maximum rate observed during the reaction following the lag period as illustrated in Figure 3-29.

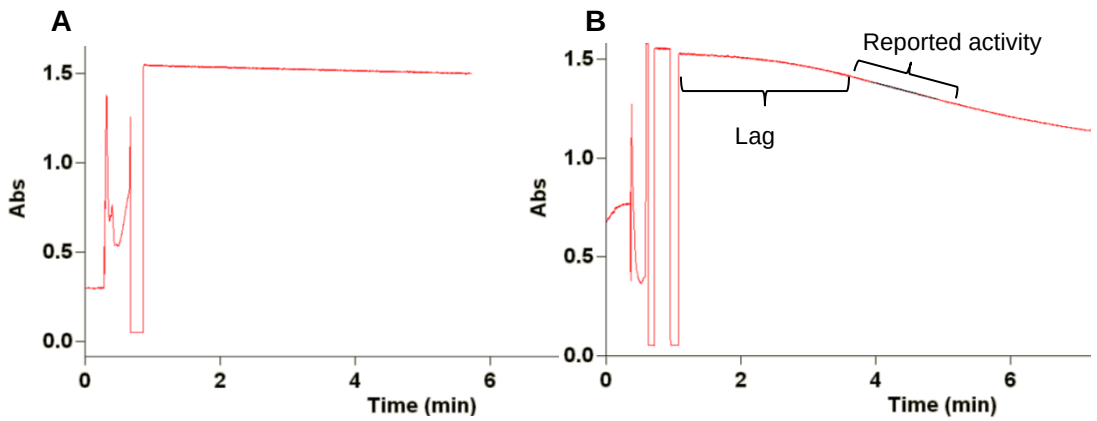


Figure 3-29: Example assay trace for the recombinant aldDH activity. A = no acetyl-CoA control, B = full assay mixture. Abs = absorbance measured at 340 nm.

Results of the aldDH activity assays (using the modified aldDH NADH assay described in Section 2.13.1.2) are shown in the following figures.

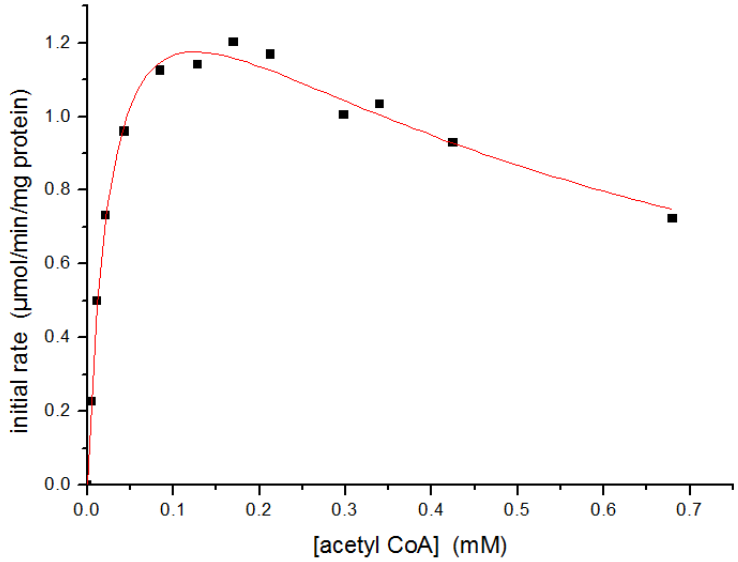


Figure 3-30: Michaelis-Menten plot of aldDH enzymic activity (U mg^{-1} of protein) against concentration of acetyl-CoA (mM) at a fixed concentration of NADH (0.22 mM) for partially-purified recombinant ADHE.

Non Michaelis-Menten kinetics were observed for the aldDH activity with respect to acetyl-CoA, these data appear to fit the trend expected for substrate inhibition.

Accurate kinetics for the aldDH domain with respect to NADH were unable to be determined. The nature of the extended assay meant that at low levels of NADH the increasing rate of reaction was more difficult to resolve.

3.4.5.2 ADH

Results of the ADH activity assays (using the ADH assay Section 2.13.2):

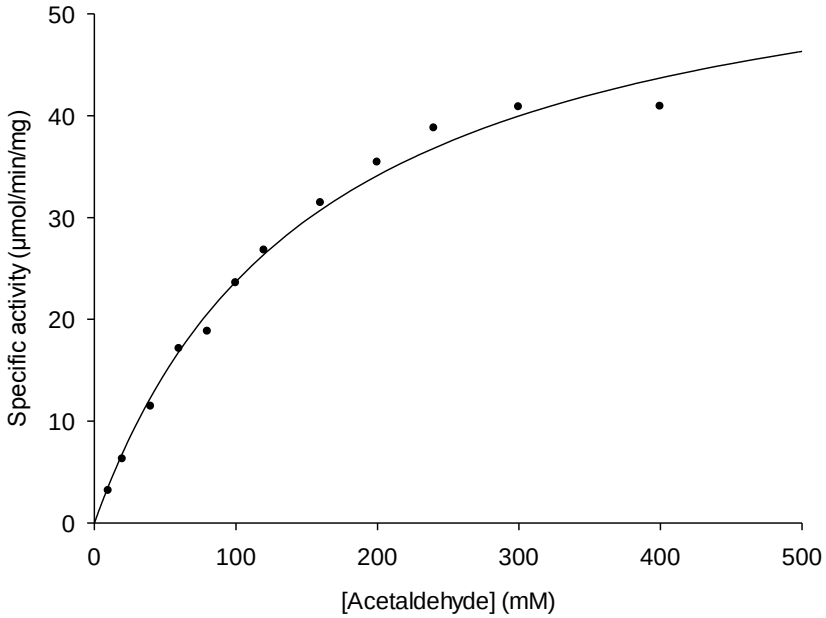


Figure 3-31: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) against concentration of acetaldehyde (mM) at a fixed concentration of NADH (0.22 mM) for partially-purified recombinant ADHE.

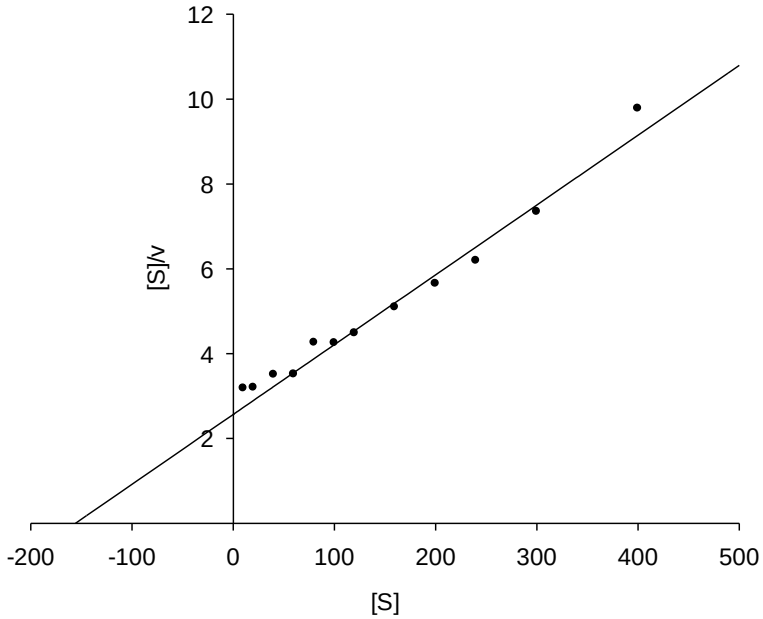


Figure 3-32: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of acetaldehyde (mM).

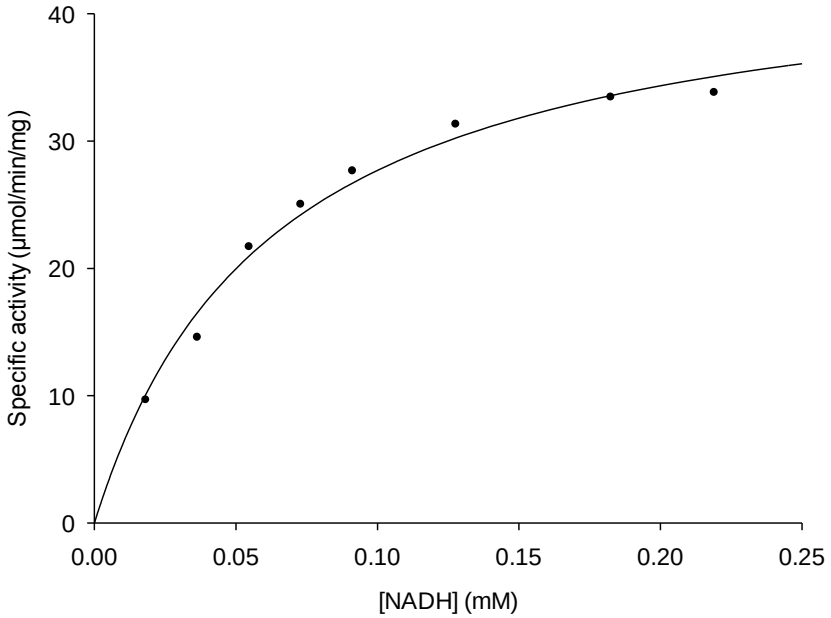


Figure 3-33: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) against concentration of NADH (mM) at a fixed concentration of acetaldehyde (240 mM) for partially-purified recombinant ADHE.

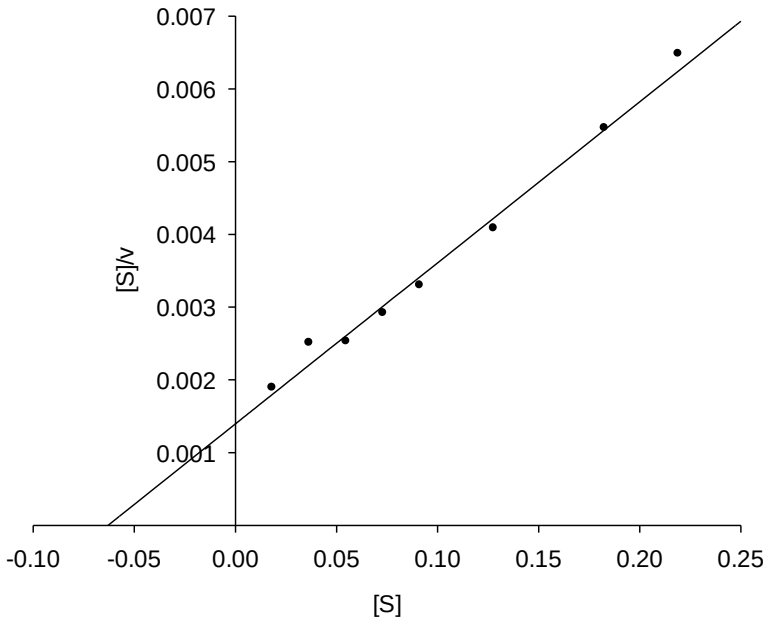


Figure 3-34: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of NADH (mM).

3.4.5.3 Summary of kinetic analyses

Activity	Substrate	Kinetic parameter	Parameter value		Units
ADH	Acetaldehyde	K_m	201.0	+/- (5.9)	mM
		$V_{max\ (adjusted)}$	91.4	+/- (5.2)	U mg^{-1}
ADH	NADH	K_m	0.065	+/- (0.004)	mM
		$V_{max\ (adjusted)}$	85.6	+/- (2.9)	U mg^{-1}
AldDH	Ac-CoA*	K_m	0.029	+/- (0.003)	mM
		V_{max}	1.7	+/- (0.08)	U mg^{-1}

Table 3-8: Kinetic parameters determined for partially-purified recombinant ADHE. *Kinetic parameters for aldDH with respect to acetyl-CoA were determined by fitting the data to the substrate inhibition equation as described in Section 2.13.7.3.

3.4.6 Optimum temperature of recombinant ADHE activity

The optimum temperature of the ADHE enzyme activities were determined by assaying the enzymes using the standard assays at varying temperatures as described in Section 2.13.5.

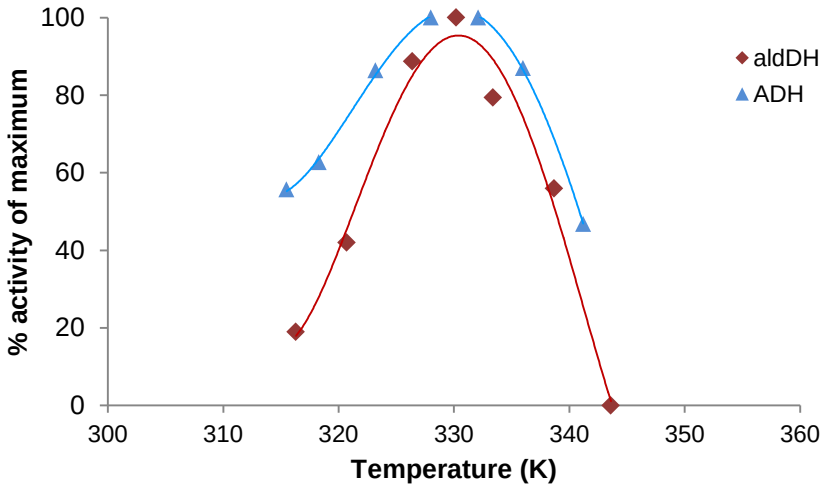


Figure 3-35: Plot of % activity of maximum against temperature (K) for both ADHE activities in partially-purified recombinant ADHE extracts.

The optimum temperature for both activities appeared to be approximately 57°C.

3.4.7 Zinc stimulation of ADH activity

When assayed in unfractionated cell extracts prior to purification, the recombinant ADHE protein was shown to be significantly stimulated by the addition of 0.1 mM zinc acetate; this stimulation was not observed after metal-affinity purification and dialysis into 0.2 M phosphate buffer pH 6.0, as shown in Table 3-9. No ADH activity was

detected in a cell extract of the non-transformed expression strain used when grown under similar conditions, with and without zinc in the assays.

Sample	Metal ion (0.1 mM)	% ADH activity (of activity with no metal ions)
Cell extract	None	100
	Zn^{2+}	237
Partially purified ADHE	None	100
	Zn^{2+}	105

Table 3-9: Relative effect of metal ions in the assay buffer on the ADH activity of the recombinant ADHE enzyme. Assays were performed as in Section 3.3.2.

3.4.8 Discussion

An N-terminal His-tagged version of the *Geobacillus thermoglucosidasius* (TM242) ADHE protein has been successfully cloned, expressed and partially-purified from *E. coli* expression strains. Using mass spectrometry analysis, the ADHE protein was confirmed to be the dominant species in protein samples partially-purified by metal-affinity chromatography. Some of the other co-eluting bands visible on the SDS-PAGE gel were also analysed using mass spectrometry.

As the M_r values for all the significant co-eluting protein bands observed on the SDS-PAGE gels were smaller than the predicted M_r of the ADHE protein, it is likely that some of the co-eluting bands are degradation products of the target protein. In the mass spectrometry analysis, sample 2 (the second most intense band) appears to correspond to part of the ADHE protein. It is hypothesised that this band corresponds to the ADH domain of the protein due to the low sequence coverage observed over the N-terminal region. A low level of Cpn60 (the chaperonin used in the Arctic express strain) also appears to co-purify with the protein of interest.

The role of Cpn60 in this expression strain is to act as an aid to protein folding, and interactions between this protein and the ADHE protein may explain the co-elution of the two proteins following nickel-affinity purification. A susceptibility within the linker region of the two domains to degradation may explain the number of co-purifying protein bands observed for both the recombinant and native ADHE proteins, although protease inhibitors (no EDTA) were present during their purification.

The recombinantly-produced ADHE protein in this section of the project has been shown to be catalytically active with both ADH and aldDH activities present in the

partially-purified protein samples. Both these activities do differ significantly from the ADHE enzyme partially isolated from TM242 cell extracts.

The aldDH activity of the recombinant protein was significantly lower than that observed for the native protein assuming equal purity (1.7 U mg^{-1} and 6.22 U mg^{-1}, respectively, for the partially-purified proteins). The loss of activity due to the predicted oxygen sensitivity of this protein observed for the native protein may also explain the lower levels of aldDH activity observed for the recombinant protein. Reducing agents were not included in the metal-affinity chromatography purification due to the problems caused by reduction of nickel ions on the column. Assays of the cell extract immediately post cell lysis also showed a relatively low aldDH compared to the ADH activity.

This relatively low specific activity observed for the aldDH domain, along with the activity lag seen in enzyme assays, suggests this domain of the protein may not be correctly/fully folded when produced using the *E. coli* expression strains used in this work. It should also be noted that the growth of the *E. coli* strains was under aerobic conditions (500 ml medium in 2.5 L baffled flasks with 250 rpm shaking), whereas the native material was from anaerobically grown cells. This may also have a significant effect on the activity/stability of the aldDH domain of ADHE. The apparent requirement of the recombinant aldDH domain of the protein for pre-incubation prior to an increased enzymatic rate may also be due to the low temperature of expression used (12°C) when producing the protein recombinantly.

The non Michaelis-Menten kinetics observed for the ADH in the native protein experiments were not observed for the recombinantly-produced protein, confirming that this effect was due to the presence of a second ADH activity. The activity observed for the ADH domain was higher than that observed for the native protein (91.4 U mg^{-1} and 51.5 U mg^{-1}, respectively, in the partially-purified samples). The difference here is less significant than that observed for the aldDH domain and is likely to be due to differences in protein purity. A significant difference in the K_m for acetaldehyde (80 mM in native, 201 mM for the recombinant) was observed between the differently produced ADHE proteins. The lack of influence of the “ADH2” protein may go some way to explaining the higher K_m observed for the recombinant ADHE. Another possibility is that if, as hypothesised, the aldDH domain of this protein is incorrectly folded, it may interfere with the ADH domain of the protein, increasing the observed value of K_m.

The optimum temperature observed for the recombinant enzyme is similar to that observed for the native protein. These results are consistent with the optimum growth temperature for the organism of approximately 60°C.

Zinc stimulation of the ADH activity of the recombinant ADHE protein in cell extracts is consistent with the stimulation observed for the native protein previously; the variation in the magnitude of this effect is likely to be due to differences in protein expression conditions. It is hypothesised that this is not observed for the recombinant protein post purification due to its scavenging Ni^{2+} ions from the metal-affinity column, thus limiting the activation by additional metal ions due to saturation. The influence of metal ions on the ADH domain of ADHE is further discussed in subsequent chapters.

During the purification of both the native and recombinant ADHE proteins it was observed that the protein consistently appeared to elute in the void volume of the gel filtration column used for purification. This provided an estimate of M_r of greater than 1.3×10^6 which would correspond to a protein complex of larger than 13 ADHE polypeptides (9.6×10^4 per ADHE). The final section of the chapter is concerned with confirming and understanding this observation.

3.5 Investigation into the multimeric assembly of ADHE proteins

3.5.1 Results

As discussed in the introduction to this chapter, ADHE has been suggested to be a protein capable of forming large multimeric assemblies known as “spirosomes”. Initial observations that ADHE appeared to elute as a large multimeric assembly from a gel filtration column, suggested this may also be the case for the ADHE protein from TM242. A series of experiments were carried out in an attempt to confirm this hypothesis.

3.5.2 Gel filtration

As described in Section 3.3.8.2, the native ADHE protein appeared to elute in the void volume of the gel filtration column used under a variety of conditions.

3.5.2.1 Recombinant ADHE

The recombinant ADHE protein was also run on the gel filtration column following His-purification:

His-purified recombinant ADHE (9 ml) was concentrated down to 2 ml using a Vivaspin 5K MWCO centrifugal filter device (Sartorius); 1 ml of this concentrated sample was loaded onto a pre-equilibrated GE Superdex 200 10/300 GL column run at a flow rate of 0.5 ml/min. The buffer consisted of 60 mM pyrophosphate pH 6.0, 0.1 M KCl, and 0.1 mM zinc acetate. Fractions were assayed to identify the peak of ADHE activity.

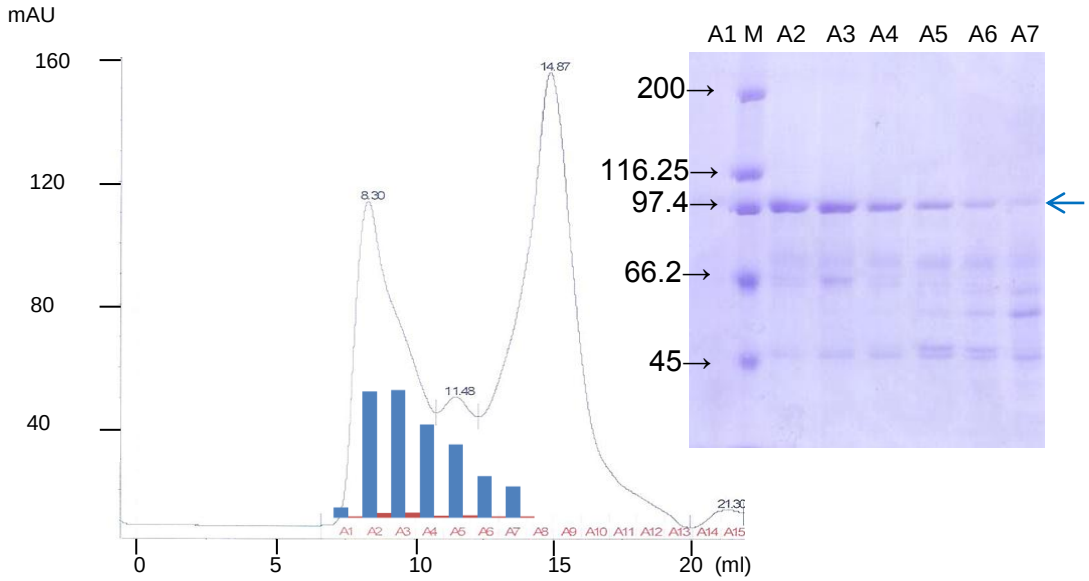


Figure 3-36: Chromatograph of a recombinant ADHE gel filtration purification. Blue line = A_{280nm} (mAU). Peak of activity measurements (µmol/min/ml) for both ADH and aldDH are overlaid onto corresponding fractions (blue = ADH red = aldDH). An SDS-PAGE gel of the peak fractions is also shown. (Fraction numbers shown on gel M = markers (M_r/1000), predicted His-tagged-ADHE band (M_r = 98,456) is highlighted with a blue arrow).

As shown in Figure 3-36, the recombinant ADHE protein also eluted in the void volume of the column.

3.5.2.2 Superose 6 gel filtration of native ADHE

The partially-purified native protein was also run on a Superose 6 10/300 GL gel filtration column (GE healthcare), which has a higher exclusion limit (4 x 10^7) than the Superdex 200 column (1.3 x 10^6). 0.6 ml of concentrated gel filtration sample was loaded onto a pre-equilibrated GE Superose 6 10/300 GL column run at a flow rate of 0.3 ml/min. The buffer consisted of 50 mM EPPS pH 8.0, 5 mM EDTA and 10% (v/v) glycerol. Fractions were run on SDS-PAGE gels to confirm the presence of the ADHE protein.

Fractions A7-11 (corresponding to the peak of ADHE) were pooled and concentrated for analysis by dynamic light scattering.

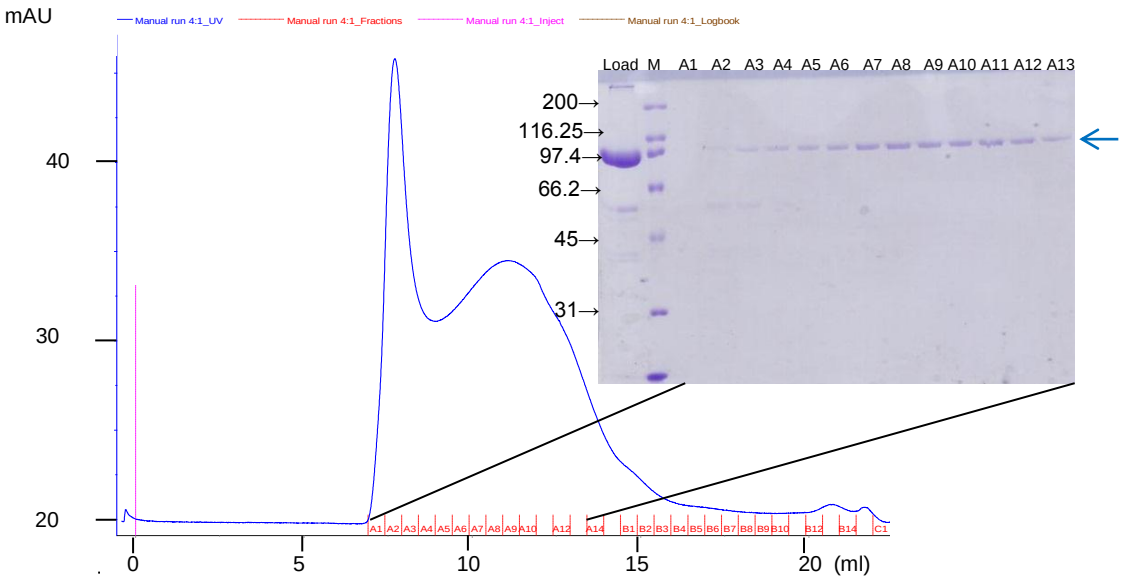


Figure 3-37: Chromatograph of a native ADHE gel filtration on Superose 6 column. Blue line = A_{280nm} (mAU). An SDS-PAGE gel of the peak fractions is also shown. (Fraction numbers shown on gel: M = markers (M_r/1000); predicted ADHE band (M_r = 96,291) is highlighted with a blue arrow).

3.5.3 Dynamic light scattering (DLS)

In an effort to determine whether the sample was mono-dispersed the concentrated pooled sample from Section 3.5.2.2 (0.048 mg/ml) was subjected to dynamic light scattering analysis as described in Section 3.2.4.

The DLS analysis showed that the purified protein sample was virtually mono-disperse; a small amount of very large particles were observed in the intensity plot but there appears to be one dominant protein species present in the sample. The globular protein estimate for this species was 4883.8 kDa. This estimate assumes that the protein is globular in nature and therefore may be inaccurate especially if the hypothesised spirosome complexes are present.

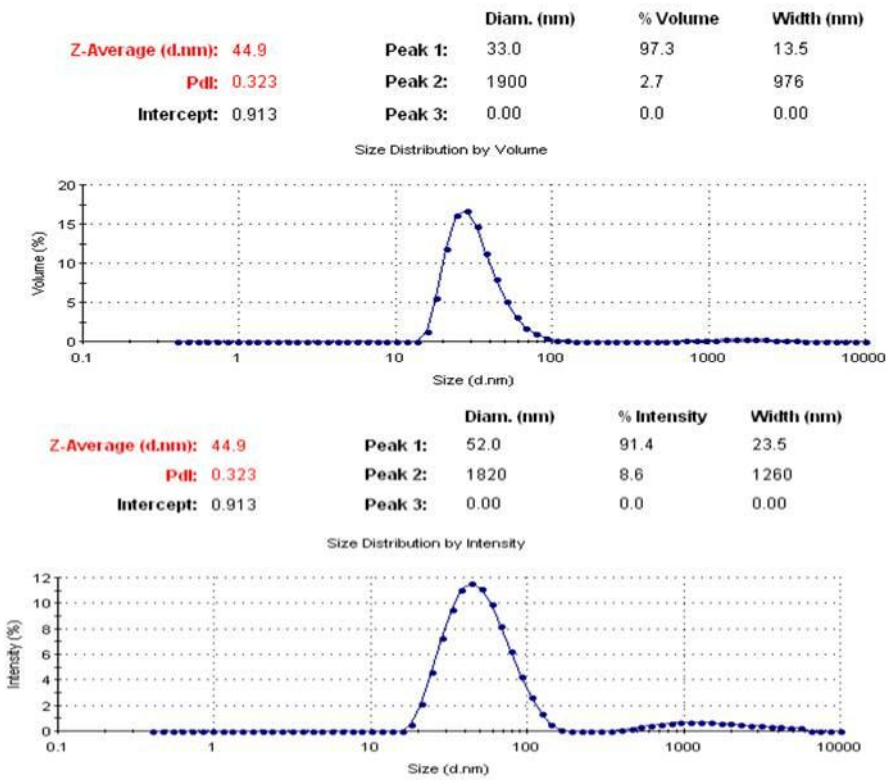


Figure 3-38: DLS analysis: Both distributions by volume and intensity results are shown; a peak value of 33 nm was used for the globular protein estimation.

3.5.4 Nanosight analysis

Samples of purified native protein (as in Section 3.5.2.2 in the presence of 5 mM GSH) and recombinant protein (as in Section 3.4.3) were also subjected to analysis using a NanoSight LM10 instrument as described in Section 3.2.5. To confirm whether any large particles observed were indeed the protein of interest, samples were compared to buffer only control samples. Both samples appeared to contain large protein particles in the 50-300 nm range. The peak of the native ADHE was approximately 200 nm, whereas that of the recombinant sample was lower at approximately 110 nm.

3.5.4.1 Native ADHE

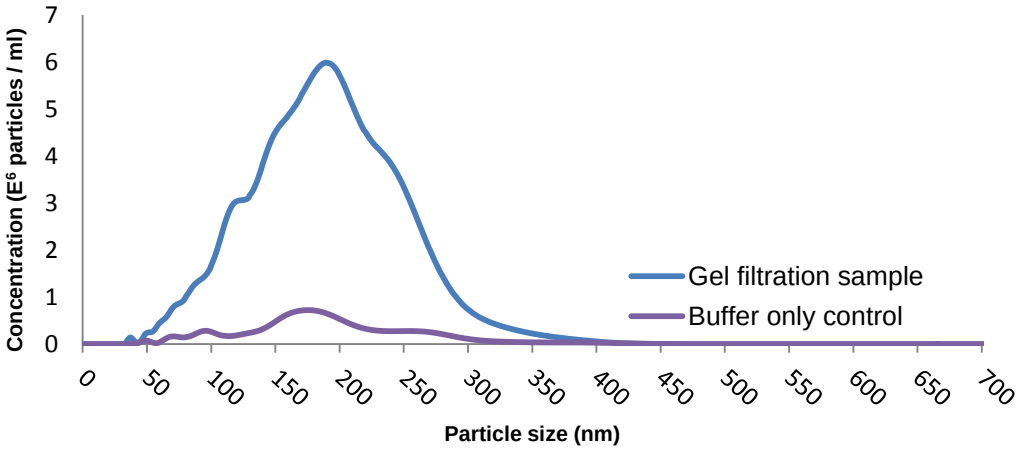


Figure 3-39: Native ADHE NanoSight particle distribution.

3.5.4.2 Recombinant ADHE

A negative control sample for the recombinant ADHE was made by growing Arctic Express© cells (no ADHE construct) in the same way as protein expression cells; extracts were produced in the same way as for protein expression i.e. metal-affinity purification. This preparation acted as an expression strain control to ensure it was the recombinant protein that was being detected.

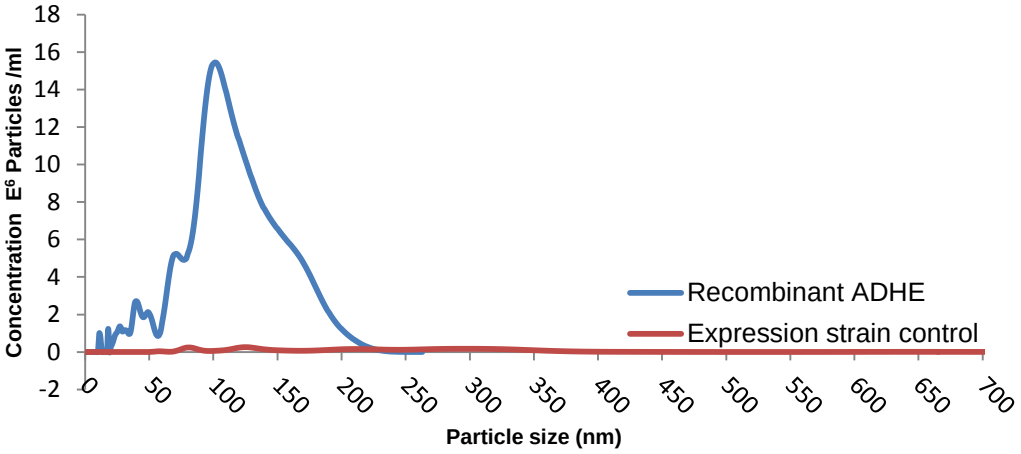


Figure 3-40: Recombinant ADHE NanoSight particle distribution.

3.5.5 Electron Microscopy

Partially-purified ADHE protein (as in Section 3.3.8.2) was used for "spirosome" (Bruchhaus and Tannich 1994; Kessler et al. 1992; Matayoshi et al. 1989) visualisation attempts using TEM (as in Section 3.2.6). Protein samples at 23, 50 and 100 µg/ml in 50 mM EPPS pH 8.0, 5 mM GSH 5 mM EDTA and 10% (v/v) glycerol were used for the experiments; 2% Uranyl-Acetate (pH 4) and 2% phosphotungstic acid (pH 8) (PTA)

were both tried as negative staining agents for the samples. The optimum contrast and minimal artefacts were observed when using PTA with the 50 µg/ml protein sample.

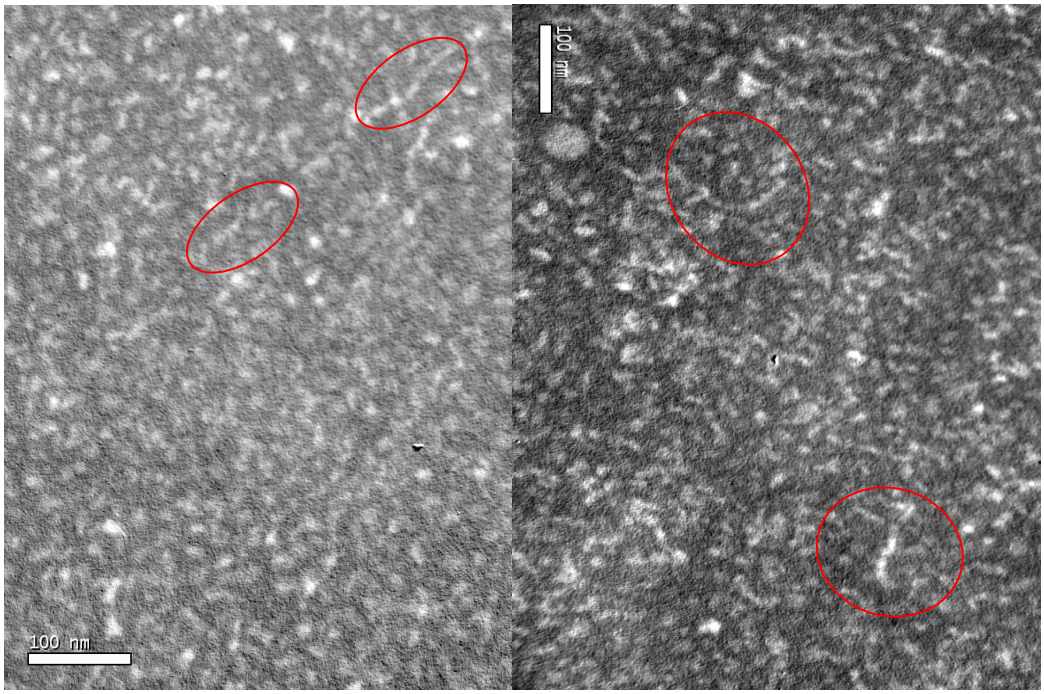


Figure 3-41: Representative TEM images for the ADHE protein samples; possible indications of spirosomes are indicated in the figure. The bar shown indicates 100 nm at the magnification used for the images.

The low resolution of the images obtained means these results are not conclusive for the presence or absence of spirosomes; there is, however, an indication of larger protein structures that appear to be linear in nature.

3.5.6 Dissociation experiments

Experiments were carried out to determine whether there is a link between the assemblies of ADHE particles and enzyme activity.

Anion exchange samples of TM242 cell extracts as in Section 3.3.8.2 were dialysed overnight into the gel filtration buffer (50 mM EPPS pH 8.0, 5 mM EDTA, 5 mM GSH, and 10% (v/v) glycerol) containing 0 M, 0.5 M and 1 M urea, prior to concentration for gel filtration. The gel filtration stage was then carried out in the standard buffer (with 0 M 0.5 M or 1 M urea) to determine if the large protein aggregates had dissociated into smaller units. In all these experiments, the characteristic void volume peak in the gel filtration run was observed (as in Section 3.3.8.2), corresponding to the large ADHE protein assemblies/aggregates suggesting that the multimeric structures had not dissociated.

A 0 M urea sample was re-run on the gel filtration column prior to overnight dialysis into the gel filtration buffer with 4 M urea. The sample was then concentrated and loaded onto the gel filtration column and run with 4 M urea in the running buffer.

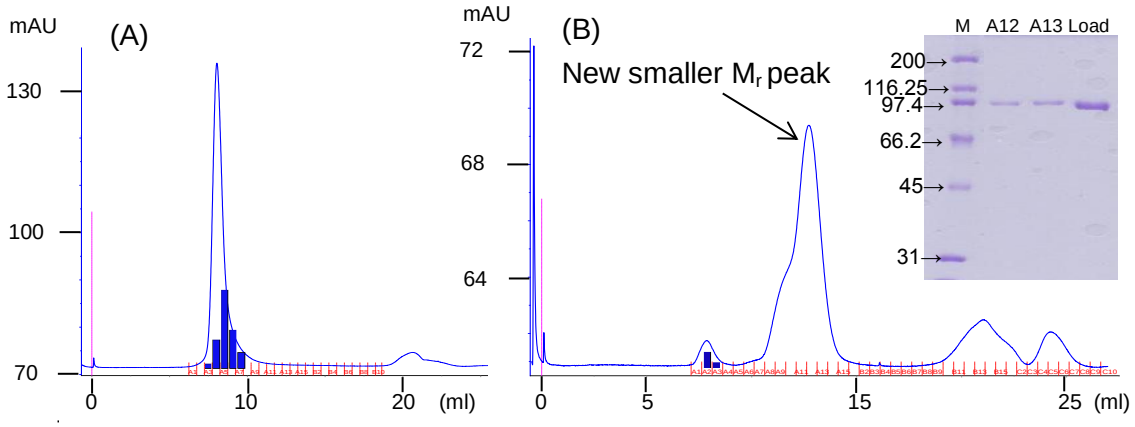


Figure 3-42: Chromatograph of the TM242 denaturant gel filtration experiment. (A) = 0 M urea and (B) = 4 M urea sample. Blue line = A_{280nm} (mAU). Peak of activity measurements (abs/min) for ADH is overlaid onto corresponding fractions. An SDS-PAGE gel of the peak fractions of the new smaller M_r protein peak is also shown (Fraction numbers shown on gel: M = markers (M_r /1000), Load = sample loaded onto column).

A second protein peak corresponding to a smaller M_r species appeared following treatment with 4 M urea. No ADH activity was detected in the new peak. A small amount of ADH activity was detected in the much reduced void volume peak.

It would appear that treatment with high concentrations of denaturant is able to cause dissociation of the ADHE particles. The successfully dissociated form of ADHE appeared to be inactivated by the denaturant treatment. Comparison of the second peak retention volume with standard proteins as described in Section 2.12 gave an estimate of molecular weight of 188,000 which corresponds to approximately two ADHE monomers (96,291).

3.6 Discussion

Some ADHE proteins have been shown to be capable of forming large multimeric assemblies known as “Spirosomes” (Bruchhaus and Tannich 1994; Kessler et al. 1992; Matayoshi et al. 1989; Nnyepi et al. 2007). The role that these multimeric assemblies play is not entirely clear but they do appear to be dynamic structures capable of responding to the presence of substrates as shown by Kessler et al. (1992).

The work carried out in this section of the project provides some evidence that the ADHE produced by TM242 may also form these large multimeric assemblies. Both the native and recombinant ADHE proteins were shown by gel filtration to form large protein assemblies or aggregates which were eluted in the void volume of the standard gel filtration column used. The Superose 6 column purification was used in an attempt to remove other contaminant proteins within the sample prior to light scattering analysis, but was not calibrated with standard proteins so could not be used to provide further estimates of M_r.

Light scattering analysis with the DLS and NanoSight techniques suggested samples contained large particles. DLS provided an estimate of molecular weight of 4883 kDa which would correspond to approximately 50 ADHE monomers (M_r/1000 = 96.3). NanoSight analysis suggested a range of particle sizes with a peak of 190 nm for the native material and 102 nm for the recombinantly expressed material. These size estimates are of a similar size to those described in the references discussed above, which suggests a range of 50-120 nm for the rod like particles (~40 monomers) (Kessler et al. 1992). As the light-scattering techniques are based around globular proteins it is not surprising that the observed particle size is larger than expected for these potentially “rod-like” protein complexes, estimated by the authors from electron microscopy images.

It would appear that the formation of these large multimeric assemblies of proteins is required for ADHE activity, as shown through dissociation experiments using urea. Dissociation of the multimeric assemblies into protein dimers coincided with the loss of activity observed. Due to the high concentration of denaturant required to dissociate the multimers, it should be noted that the tertiary structure of the enzyme itself may have been adversely affected by the denaturant, thus causing a loss of activity. Further investigation is therefore required to determine if the multimeric assemblies are truly required for enzyme function.

It is unfortunate that the images obtained using TEM do not provide high enough resolution to confirm the presence or absence of spirosomes in the samples.

3.7 ADHE general discussion

The characterisation of the ADHE protein from *Geobacillus thermoglucosidasius* has been described when produced from native material and recombinant expression strains. This bi-functional protein possesses both aldDH and ADH activities; it remains

unclear from these experiments as to whether both or only one of the activities present on the protein are required for ethanol production.

The aldDH domain of the protein was shown to be relatively unstable compared to the ADH domain, especially in terms of oxygen sensitivity. The aldDH domain of ADHE has also been suggested to be unstable when purified from other organisms (Pei et al. 2010). Kinetic characterisation of this domain using the NADH based assay suggests it may be under the influence of substrate inhibition with respect to acetyl-CoA. Work carried out later in the project using the DTNB assay at pH 6.5, did not appear to be subject to this substrate inhibition effect. This may therefore be an artefact of the assay conditions used rather than a property of the enzyme itself. The aldDH assay was shown to be acting as a measure of metabolic flux to ethanol (2 NADH converted per CoA-SH released) so the substrate inhibition observed may be due to a limited rate of channelling of the acetaldehyde to the ADH domain of the protein.

Estimates of the catalytic efficiencies of the two domains of the ADHE protein are outlined in the following summary table:

Activity	Substrate	V_{max}^{app} (U mg^{-1})	K_m (mM)	V_{max}^{app}/K_m ((U mg^{-1})/mM)
TM242 Native ADHE partially-purified				
ADH	Acetaldehyde	51.5	79.5	0.7
AldDH	Ac-CoA	6.22	0.006	1065.1
Recombinant ADHE partially-purified				
ADH	Acetaldehyde	70.5	201	0.4
AldDH	Ac-CoA	1.7	0.029	59.3
TM242 Native "ADH2" partially-purified				
"ADH2"	Acetaldehyde	4.26	0.184	23.2

Table 3-10: Summary of catalytic efficiencies for ADHE from native material and partially-purified recombinant material.

The ratio k_{cat}/K_m (or V_{max}/K_m) is conventionally used to provide a measure of the catalytic efficiency of an enzyme for a variety of different substrates. However, it should be used with considerable caution to compare two enzymes acting on the same substrate (Eisenthal et al. 2007; Fox and Clay 2009) or, as in Table 3-10, to compare the efficiency of, and flux through, two enzymes acting in sequence. As this measure of catalytic efficiency is determined at high concentrations of substrate, the *in vivo* implications of the derived parameter is limited. The k_{cat}/K_m (or V_{max}/K_m) value

determined for two enzymes may be identical, but the rates observed at different substrate concentrations may differ substantially due to relative differences in the k_{cat} and K_m values (Cornish-Bowden and Cardenas 2010). Without determining the intracellular concentrations of the substrates involved, more suitable measures such as catalytic competence (Ceccarelli et al. 2008) also cannot be used.

Considering the native ADHE, it would appear that the aldDH is significantly more efficient catalytically than the ADH enzyme, even though the V_{max} for ADH is 8-fold higher. However, the *in vivo* flux through the two enzymes will depend on the intra-cellular concentration of acetyl-CoA and NADH with respect to their K_m values, and the very high K_m of ADH for acetaldehyde may well be overcome by substrate channelling between the two enzymes. A similar argument may apply to the recombinant ADHE, where again there is a significant difference in the values of V_{max}/K_m for the aldDH and ADH enzymes. The over-estimation of aldDH rates in the NADH based aldDH assays may also artificially increase the difference between the two activities.

Of the work carried out on ADHE enzymes from other organisms, a majority has focussed on confirming the presence and activity of the protein rather than full kinetic characterisation (Chen et al. 2004; Fontaine et al. 2002; Nair et al. 1994; Peng et al. 2008). Specific activities under defined conditions are commonly quoted, but specific kinetic parameters are not. Of those that do quote kinetic parameters, some of the assays have been carried out in the non-physiological direction (Membrillo-Hernandez et al. 2000). Bruchhaus et al (1994) and Espinosa (2001) determined the kinetic parameters for the enzyme from *E. histolytica* in the physiological direction (Bruchhaus and Tannich 1994; Espinosa 2001). Bruchhaus et al (1994) determined the K_m values for the ADH to be 0.15 mM (acetaldehyde) and 0.05 mM (NADH), whereas the aldDH values are 0.015 mM (acetyl-CoA) and 0.18 mM (NADH) (comparable values were obtained by Espinosa et al (2001)). The relative activities stated for partially-purified protein were ADH 233 U mg^{-1} and aldDH 87 U mg^{-1}. The K_m value obtained for acetaldehyde in this project for the ADH domain of the protein is significantly higher than that for the *E. histolytica* enzyme. This may be due to other effects not taken into account such as oxygen sensitivity, protein mis-folding, differences between the proteins or in the assay conditions used.

Substrate channelling between the two domains may explain the relatively high K_m values determined for acetaldehyde for the ADH domain of the protein. Substrate channelling can be defined as the passing of the product of one enzyme within a

pathway as the substrate to the subsequent enzyme without going via the bulk solution. This phenomena is common within metabolic pathways and is discussed in a recent review (Zhang 2011). Many factors have been suggested to be involved in channelling, including the proximity of the active sites of the two enzyme activities. This may be the case for the ADHE protein with two distinct domains held together by a linker region between them. A particularly important aspect of channelling may be when the product of one enzyme may have toxic effects on the organism in which it is produced. Acetaldehyde is a reactive compound that is likely to have a negative effect on the cell if high concentrations were allowed to accumulate, and so efficient channelling of this compound may be favourable. If this were the case for ADHE, the substrate would be released by the aldDH close to the ADH active site. The ADH would therefore not tend to encounter acetaldehyde free in solution. Assaying the enzyme by adding acetaldehyde to the solution may therefore not be a fair reflection of the conditions this domain would normally experience within the cell.

Multimerisation of proteins has also been suggested to be linked to channelling (Perham 1975; Schoffelen and van Hest 2011). The likely formation of large multimeric protein assemblies (spirosomes) shown here can be envisaged to play a part in the efficient channelling of the potentially toxic intermediate product of the ADHE enzyme.

It is conceivable that it is only the aldDH domain of ADHE that is essential for ethanol production in TM242. The presence of a second ADH activity in the TM242 cell extract (when grown anaerobically) and the high K_m of the ADHE ADH domain support this hypothesis. Work in the following chapters attempting to resolve the two domains of ADHE was carried out with the intention of using any active truncated aldDH fragments to test this hypothesis.

The competition between the aldDH of ADHE and the phosphate acetyl-transferase (PAT) for acetyl-CoA in TM242 is a key factor in maximising ethanol production by this organism (Figure 1-3). The activity of PAT is being investigated in another PhD CASE studentship in conjunction with TMO Renewables; preliminary data suggest that PAT may be a more catalytically active enzyme than the aldDH domain of ADHE (Hills, C 2011 unpublished work). Other aldDH proteins are annotated within the *G. thermoglucosidasius* genome; later parts of this project were focused on identifying a more stable/catalytically active aldDH activity from these proteins.

It would appear that the ADH activity of ADHE is positively affected by the presence of various divalent metal ions. The influence of these metal ions is further investigated in subsequent chapters of this thesis.

In conclusion, the ADHE enzyme from TM242 has been characterised and shown to possess potential limitations under the conditions used here, in terms of aldDH stability and ADH catalytic efficiency. The presence of a second ADH activity within cell extracts of TM242 may overcome the limitations of the ADH domain of ADHE. The protein appears to form large multimeric assemblies as observed for other ADHE proteins, but the purpose of these assemblies is currently unresolved.

4 ISOLATION AND CHARACTERISATION OF INDEPENDENT ADHE DOMAINS

4.1 Introduction

The structural interdependence of the two domains of ADHE has been investigated in *E. histolytica* through a series of papers by Stanley's group. Their work has shown that the two domains of the ADHE protein appear to be structurally independent and are able to be expressed individually (Chen et al. 2004; Espinosa 2001). The 870 amino acid protein was successfully resolved into two functional protein fragments. The aldDH activity was found on a fragment containing amino acids 1-446 and the ADH activity was isolated on a fragment containing amino acids 417-870. In the earlier paper (Espinosa 2001), aldDH activity was not detected on a fragment containing amino acids 1-532; the authors suggested that interference from C-terminal residues may be responsible for the lack of activity in this fragment.

The ADH fragment obtained had a significantly lower K_m for acetaldehyde and NADH than that of the full length protein, suggesting that removing this domain from ADHE may affect the kinetic properties of the enzyme. Given the high value of K_m obtained for the ADH domain for the protein in Chapter 3, it was hypothesised that this may also be the case for the protein from *G. thermoglucosidasius.*

Alignment scores between the two ADHE proteins show 57% identity with 74% similarity (2% gaps). The experiments carried out here were designed to replicate the fragmentation carried out for the *E. histolytica* ADHE in *G. thermoglucosidasius,* attempting to obtain the minimal functional units required for both aldDH and ADH activity from this protein.

The library of protein fragments designed for these experiments is shown in Figure 4-1. Fragments were designed around the hypothetical motifs annotated within the ADHE sequence. These motifs were identified using the PROSITE tool (ExPASy) (Sigrist et al. 2010) using the tool's database and searching for known NAD^+ binding motifs.

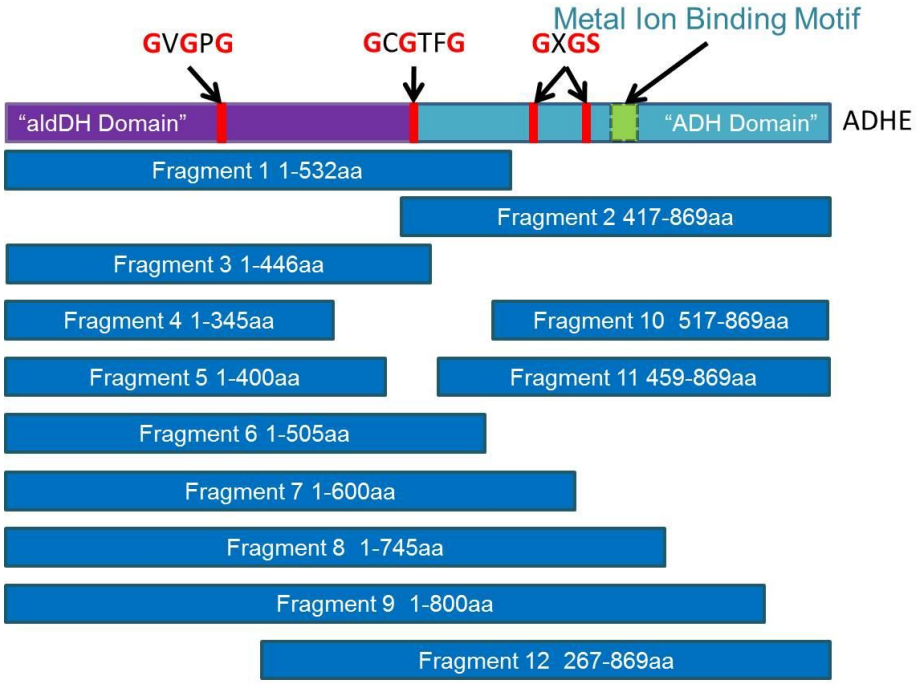


Figure 4-1: Overview of fragmentation of ADHE (scale diagram). Red blocks indicate hypothetical NAD^+ binding sites. The predicted metal ion binding site (AIVDPQFVMTVPKHVTADTGMDVLTHAIE) is also indicated.

The TM400 and TM393 strains were initially used to express the fragments due to the potential mis-folding of the aldDH domain in *E. coli.* Protein fragments were cloned and expressed using the pUCG18-pLDH-*pheB* construct, a modified version of the vector pUCG18 (Taylor et al. 2008). This modified vector was developed by Bartosiak-Jentys (Bartosiak-Jentys 2010) specifically for expressing proteins in *Geobacillus* spp. The vector contains the LDH promoter from *G. stearothermophilus* regulating a catechol reductase marker gene *pheB*. The *pheB* gene was excised using the restriction enzymes XbaI/SacI and the fragments of interest were cloned in at this point using these sites.

Aerobic and micro-aerophilic expression conditions were used for these experiments, as expression under the LDH promoter is affected by the level of aeration of a culture (Bartosiak-Jentys 2010). The possibility that active aldDH fragments may be toxic under fermentative conditions, due to the accumulation of acetaldehyde, was also a consideration.

A selection of fragments was chosen for expression in *E. coli,* incorporating an N-terminal His-Tag into the proteins to facilitate protein purification. The pET28a vector and conventional *E. coli* expression strains were used for these experiments.

4.2 Materials and methods

4.2.1 ADHE fragment cloning

The pTMO259 construct purified from JM109 cells was used as the DNA template for the PCRs for construction of the fragments described here. These were carried out according to the method described in Section 2.3.2.

The primers used to amplify the required protein fragments with the XbaI/SacI restriction sites are shown in Table 4-1. The DNA sequences of the primers used are reported in Appendix 1. AldDH fragment reverse primers incorporated a stop codon immediately prior to the SacI restriction site. An artificial start codon was introduced immediately following the XbaI restriction site where required for the ADH fragments.

Protein fragment	Forward primer	Reverse primer
Fragment 1 (1-532 aa)	GB ADHE Fwd1	GB aldDH Rev1
Fragment 2 (417-869 aa)	GB ADH Fwd1	GB ADHE Rev1
Fragment 3 (1-446 aa)	GB ADHE Fwd1	GB aldDHshort Rev1
Fragment 4 (1-345 aa)	GB ADHE Fwd1	GB 1-345aa Rev1
Fragment 5 (1-400 aa)	GB ADHE Fwd1	GB 1-400aa Rev1
Fragment 6 (1-505 aa)	GB ADHE Fwd1	GB 1-505aa Rev1
Fragment 7 (1-600 aa)	GB ADHE Fwd1	GB 1-600aa Rev1
Fragment 8 (1-745 aa)	GB ADHE Fwd1	GB 1-745aa Rev1
Fragment 9 (1-800 aa)	GB ADHE Fwd1	GB 1-800aa Rev1
Fragment 10 (517-869 aa)	GB 517-869aa Fwd1	GB ADHE Rev1
Fragment 11 (459-869 aa)	GB 459-869aa Fwd1	GB ADHE Rev1
Fragment 12 (267-869 aa)	GB 267-869aa Fwd1	GB ADHE Rev1
ADHE (1-869 aa)	GB ADHE Fwd1	GB ADHE Rev1

Table 4-1: PCR primers used for initial fragmentation of the ADHE protein.

The PCR products for the library of fragments constructed were A-tailed, ligated into the pGEM®-T easy vector, ethanol-precipitated and transformed into JM109 cells for blue/white screening. Successful cloning of the fragments was confirmed by DNA sequencing. The fragments in pGEM®-T easy were then digested with XbaI/SacI, gel purified and ligated into the previously-digested *Geobacillus* protein expression vector pUCG18-pLDH (replacing the marker gene *pheB*). Ligations were transformed into JM109 cells, which were then subject to PCR colony screening. Positive plasmids from the PCR screen were screened by restriction digestion with XbaI/SacI, prior to positive constructs being sent for DNA sequencing with vector-specific primers. The pUCG18-pLDH-fragment vectors were transformed into electrocompetent TM400 and TM393 for protein expression trials.

4.2.2 Aerobic *Geobacillus* expression

Aerobic protein expression trials used the shake flask method described in Section 2.5.2 at 58°C.

4.2.3 Micro-aerophilic *Geobacillus* expression

The following method was used to produce micro-aerophilic protein expression cultures of *G. thermoglucosidasius*:

- A loop-full of biomass from the required glycerol stock was inoculated into 5 ml of pre-warmed (55°C) TTY medium containing 12 µg/ml Kan and allowed to grow for 2 h shaking at 220 rpm.
- This culture was then transferred into 50 ml of pre-warmed (55°C) TTY medium containing 12 µg/ml Kan in a 250 ml baffled flask and grown for 16 h shaking at 220 rpm.
- A 500 ml Duran bottle containing 550 ml of pre-warmed (55°C) TTY medium containing 12 µg/ml Kan, was inoculated with 10 ml of the overnight culture and grown for the required time period with shaking at 200 rpm.
- Cell pellets produced by this method were obtained by centrifugation at 5300 x *g* for 20 min (4°C).
- The supernatant was discarded and the pellets stored at -20°C until required.

4.2.4 Partial purification and sizing of active ADH fragments

The pI value's for the active ADH fragments were determined using the ProtParam tool to be:

- Fragment 2 pH 8.73
- Fragment 11 pH 7.72
- Fragment 12 pH 7.95

Cell extracts (prepared as described in Section 2.6, but replacing 50 mM EPPS pH 8.0 with ion-exchange buffer A) were loaded onto two pre-equilibrated ion-exchange columns (GE HiTrap 5 ml Q-Sepharose SP for Fragment 2, and GE HiTrap 5 ml Q-Sepharose HP columns for Fragments 11 and 12) and run in series at a flow rate of 1 ml/min. Proteins were eluted from the columns using a 0-1 M NaCl gradient (in buffer A) over a 60 min period. Fractions were assayed to identify the peak of ADH activity prior to pooling the highest activity fractions ready for sizing by gel filtration.

Buffer A consisted of 50 mM EPPS, 0.1 mM zinc acetate and 5 mM reduced glutathione. For Fragment 2, the buffer was pH 7.3 and for Fragments 11 and 12 pH 8.8 was used.

Prior to loading onto a GE Superdex 200 10/300 GL gel filtration column, protein samples were concentrated using a Vivaspin 5K MWCO centrifugal filter device (Sartorius) and filtered using a Ultrafree centrifugal filter device (0.45 µm) (Millipore). The column was run at a flow rate of 0.4 ml/min. The buffer consisted of 50 mM EPPS pH 8.0, 5 mM reduced glutathione, 5 mM EDTA, and 10% (v/v) glycerol. The elution profile of the protein was detected by A_{280nm} and those fractions over which the protein eluted were collected and stored on ice. Fractions were assayed to identify the peak of ADH activity prior to SDS-PAGE analysis to confirm the presence of the expected fragment.

4.2.5 His-tagged fragment cloning

Fragments 1, 2, 3 and 11 were selected for recombinant expression in *E. coli,* using the pET28a vector incorporating an N-terminal His-Tag into each fragment.

Protein fragment	Forward primer	Reverse primer	Restriction sites used
Fragment 1 (1-532 aa)	Fragment 1 (AldDH) F	Fragment 1 (AldDH) R	NheI/XhoI
Fragment 2 (417-869 aa)	Fragment 2 (ADH) F	Fragment 2 (ADH) R	NdeI/XhoI
Fragment 3 (1-446 aa)	Fragment 3 (AldDHs) F	Fragment 3 (AldDHs) R	NheI/XhoI
Fragment 11 (459-869 aa)	Frag11pET28aF	Frag11pET28aR	NdeI/XhoI

Table 4-2: PCR primers used for amplification of fragments of the ADHE protein selected for recombinant expression using pET28a.

Due to an oversight made during primer design for Fragment 1 (discovered late in the experimental phase of the project), no stop codon was introduced using the reverse primer. The next in-frame stop codon in this vector was 33 base pairs later on, meaning an extra amino acid sequence (APPPPPLRSGC-) was present at the C-terminus of this fragment.

4.3 *Geobacillus* expression hosts results

4.3.1 Fragment cloning

PCR amplified DNA fragments were sized as shown in Figure 4-2; DNA bands of the expected size were excised and gel purified. The fragments were A-tailed and ligated into the pGEM®-T easy vector overnight.

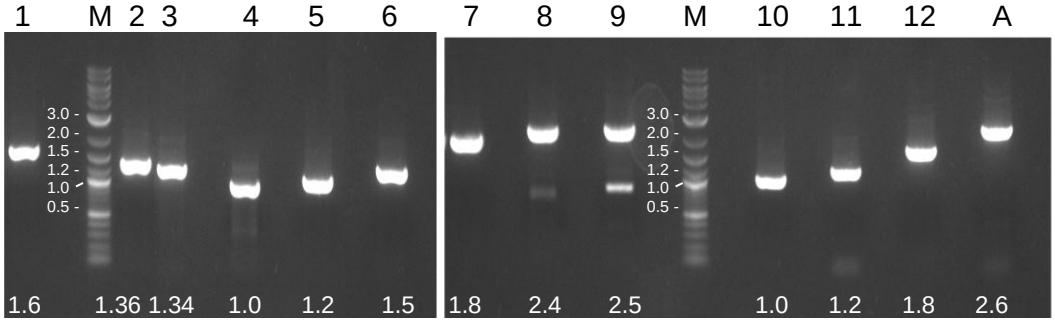


Figure 4-2: Agarose gel electrophoresis of PCR products. Lanes 1-12 = fragments 1-12, A = ADHE, M = DNA markers sizes given in kb. Predicted sizes (kb) of DNA fragments are shown below the appropriate lane.

Ligations were ethanol-precipitated and transformed into JM109 cells for blue/white screening. Isolated plasmids were purified from overnight cultures of white colonies selected for each fragment. These plasmids were screened using restriction digestion with the appropriate enzymes, and selected positive clones were sent for sequencing. Once the fragments were confirmed to be successfully cloned into pGEM®-T easy, the DNA fragments were excised using XbaI/SacI and ligated into the previously-digested and SAP-treated pUCG18-pLDH vector.

Ligations were ethanol-precipitated prior to transformation into JM109 and screened using the Carb resistance marker for *E. coli* present in this vector. Representative colonies for each fragment were picked for overnight cultures prior to plasmid preparation. Initially, the vectors were screened using restriction digestion (Figure 4-3), and positive clones were confirmed by DNA sequencing of the fragments using vector specific primers (and internal primers where required).

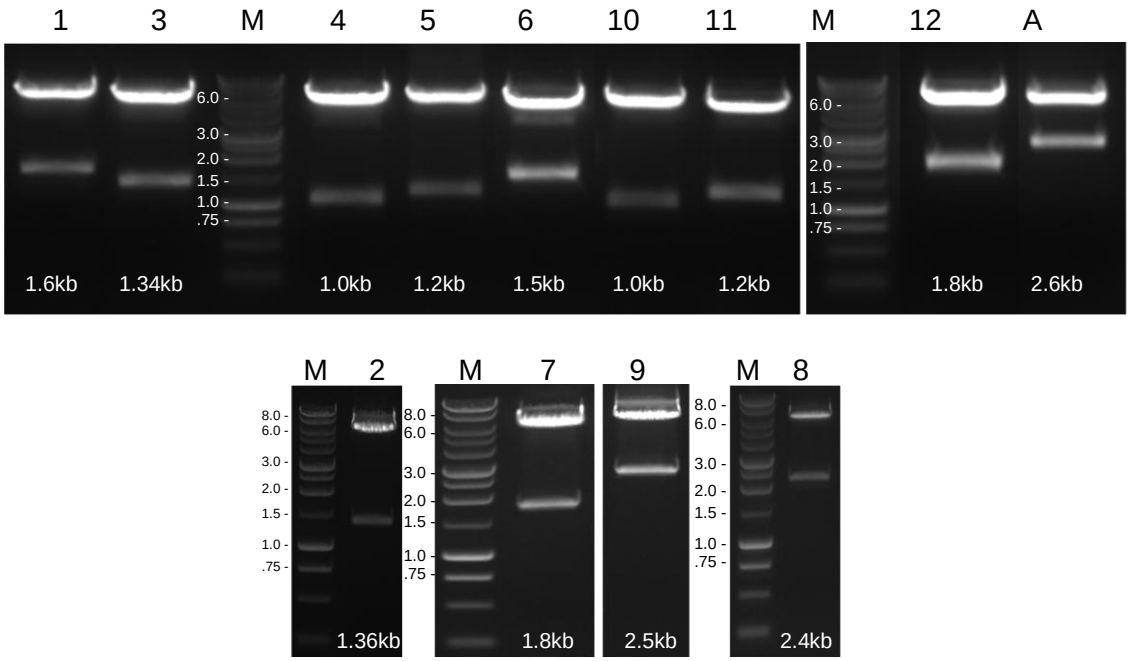


Figure 4-3: Agarose gel electrophoresis of restriction digested pUCG18-pLDH constructs. Numbers above = Fragments 1-12, A = ADHE, M = DNA markers sizes given in kb. The predicted sizes of the fragments are shown below the respective lane for each fragment. The predicted size of the pUCG-18-pLDH vector is 6.5 kb.

The purified constructs were transformed into TM400 and TM393 electrocompetent cells; screening used the Kan resistance marker for *Geobacillus* spp present in this vector.

4.3.2 Fragment expression trials (*Geobacillus*)

As discussed in Chapter 3, a low level of ADH activity was detected in cell extracts of the TM400 strain without the expression vector present; therefore, only ADH activities significantly higher than this background activity are reported here.

4.3.2.1 Aerobic TM400 trial

Cell extracts for each of the aerobically-grown TM400-fragment strains were produced, and the soluble fractions were assayed for enzymatic activities (Figure 4-4) and analysed by SDS-PAGE (Figure 4-5). The protein concentration for any soluble fraction displaying enzymatic activity was determined by the Bradford assay. AldDH activity was only detected in the cell extract from TM400-ADHE; even then, this activity was very low (0.06 U mg^{-1}). ADH activity was detected in 4 of the cell extracts tested. Fragment 11 displayed the highest activity and corresponds to the smallest functional ADH fragment produced.

The low level of aldDH activity observed in the aerobically-grown TM400-ADHE strain, is likely to be due to the oxygen sensitivity of the aldDH domain. To confirm this hypothesis and to try to improve the level of expression of the fragments, the experiment was repeated in anaerobically-grown strains.

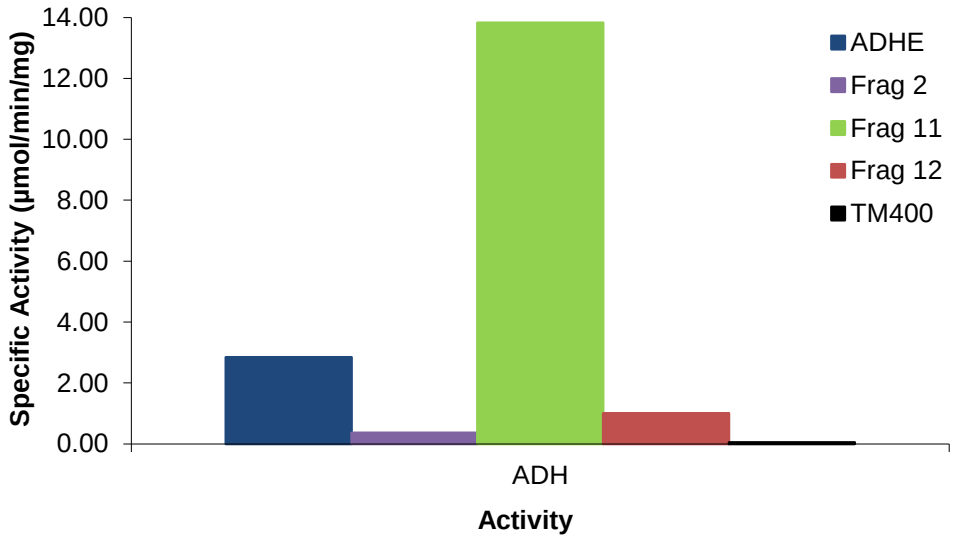


Figure 4-4: Assay data for TM400 fragments in the aerobic expression trial. Only ADH data for cell extracts displaying activity are shown.

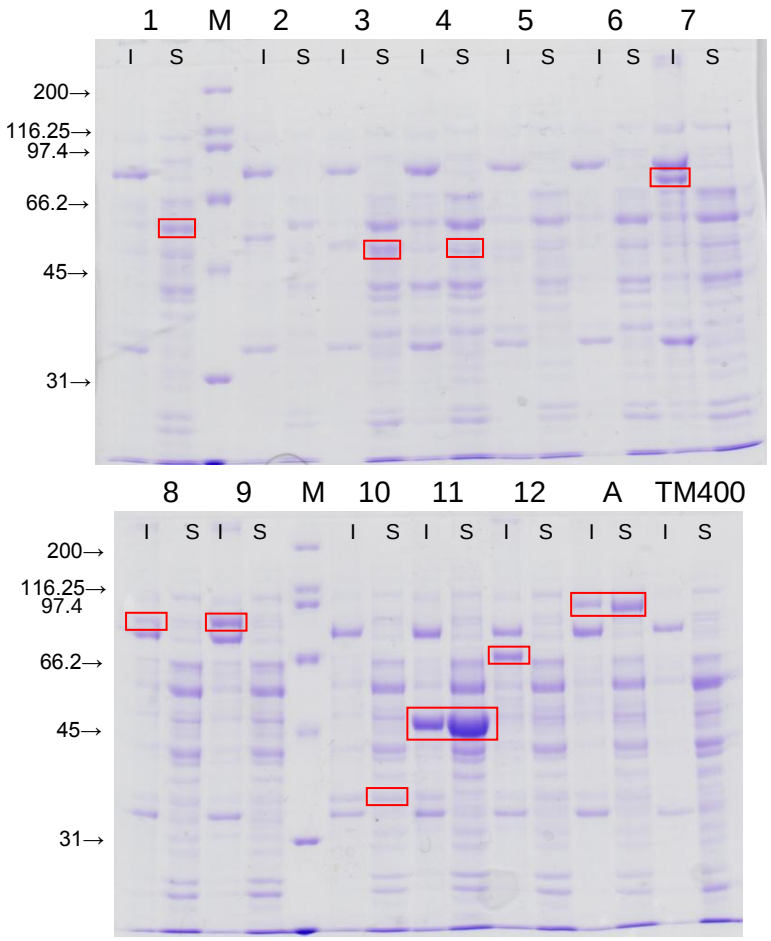


Figure 4-5: SDS-PAGE gel analysis of the aerobic TM400 expression trial samples. A = ADHE, 1-12 = fragment numbers, S = soluble fraction, I = insoluble fraction, M = markers (M_r/1000). Red boxes indicate predicted M_r of fragments and possible corresponding bands.

4.3.2.2 Anaerobic TM400 trial

Cell extracts for each TM400 strain grown for 4 h anaerobically were produced. The soluble fractions were assayed for enzymatic activities and analysed by SDS-PAGE. The protein concentration for any soluble fractions displaying enzymatic activity was determined by the Bradford assay.

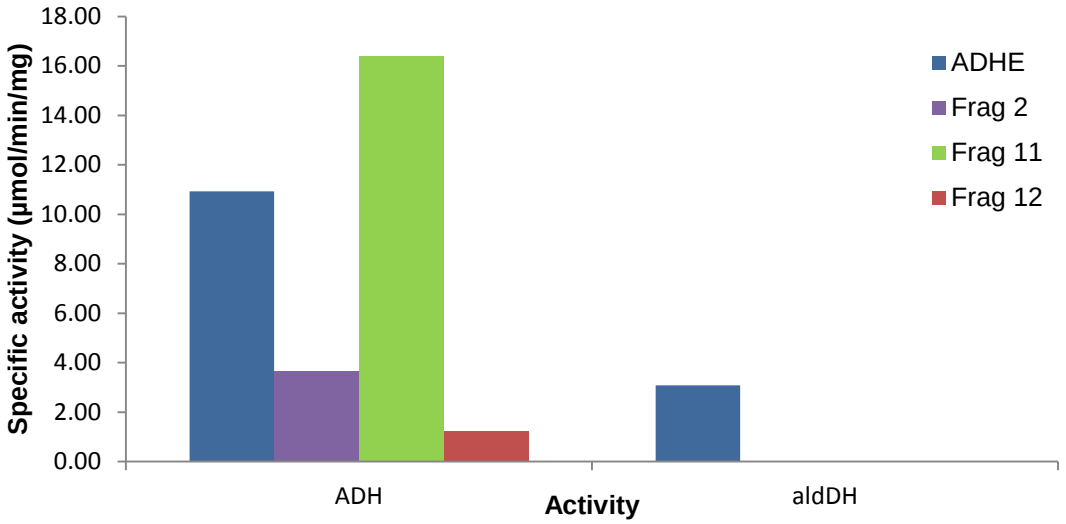


Figure 4-6: Assay data for TM400 fragments anaerobic expression trial. Only data for active extracts are shown.

AldDH activity was only detected in the cell extract from TM400-ADHE. ADH activity was detected in 4 of the cell extracts with the highest observed being for TM400-Fragment 11.

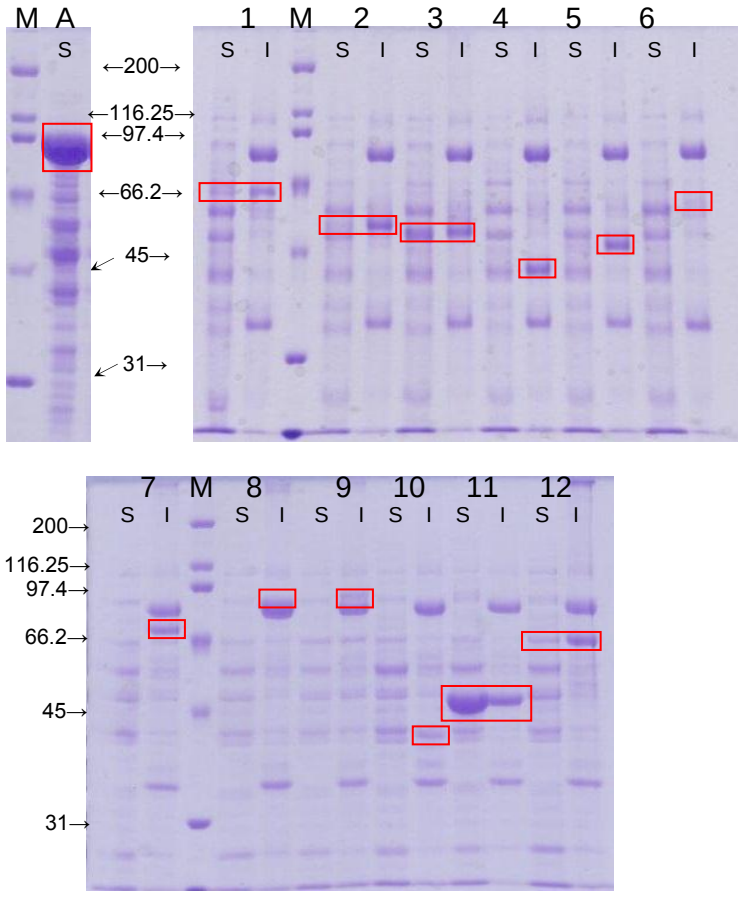


Figure 4-7: SDS-PAGE gel analysis of the anaerobic TM400 expression trial samples. A = ADHE, 1-12 = fragment numbers, S = soluble fraction, I = insoluble fraction, M = markers (M_r/1000). Red boxes indicate predicted M_r of fragments and possible corresponding bands.

Protein Identity	ADH specific activity (U mg^{-1})	AldDH specific activity (U mg^{-1})
TM400 Aerobic expression		
Fragment 2	0.4	n/a
Fragment 11	13.8	n/a
Fragment 12	1.0	n/a
ADHE	2.8	0.06
TM400 Anaerobic expression		
Fragment 2	3.7	n/a
Fragment 11	16.4	n/a
Fragment 12	1.3	n/a
ADHE	10.9	3.1

Table 4-3: Summary of specific activity measurements for catalytically-active proteins expressed in TM400 grown both aerobically and anaerobically.

Activities of all the active proteins appeared higher under anaerobic expression conditions. In both aerobic and anaerobic expression trials, TM400 Fragment 11 was the most active ADH containing cell extract. The only significant aldDH activity was detected in the anaerobically expressed TM400-ADHE. Several of the fragments were produced in the insoluble fraction of the cell extract, whilst others were poorly expressed. Similar results were obtained for expression of the fragments in the TM393 strain, with no detectable aldDH activity being observed.

4.3.3 Kinetic characterisation of the ADH domain

As the minimal functional unit of the ADH domain of ADHE, Fragment 11 was characterised kinetically in cell extracts of aerobically-grown TM400-Fragment 11 (using the overnight shake-flask method). This method was used as the anaerobic expression system produced very limited amounts of biomass. Cell extracts were diluted 1 in 10 with the standard cell extract buffer ready for assays.

During the initial protein expression trials, the level of ADH activity in aerobically grown TM400 alone was very low (0.05 U mg^{-1}) compared to Fragment 11 (13.8 U mg^{-1}), reinforcing the case for the predicted anaerobic expression of the “ADH2”. Dilution of the cell extract coupled with this relatively low activity meant that no ADH2 activity was detected in these experiments.

Activity	Substrate	Kinetic parameter	Parameter value		Units
ADH	Acetaldehyde	K_m	91.4	+/- (2.9)	mM
		$V_{max\ (adjusted)}$	28.3	+/- (1.3)	U mg^{-1}
ADH	NADH	K_m	0.04	+/- (0.001)	mM
		$V_{max\ (adjusted)}$	27.3	+/- (0.9)	U mg^{-1}

Table 4-4: Kinetic parameters of Fragment 11 in TM400-Fragment 11 cell extracts.

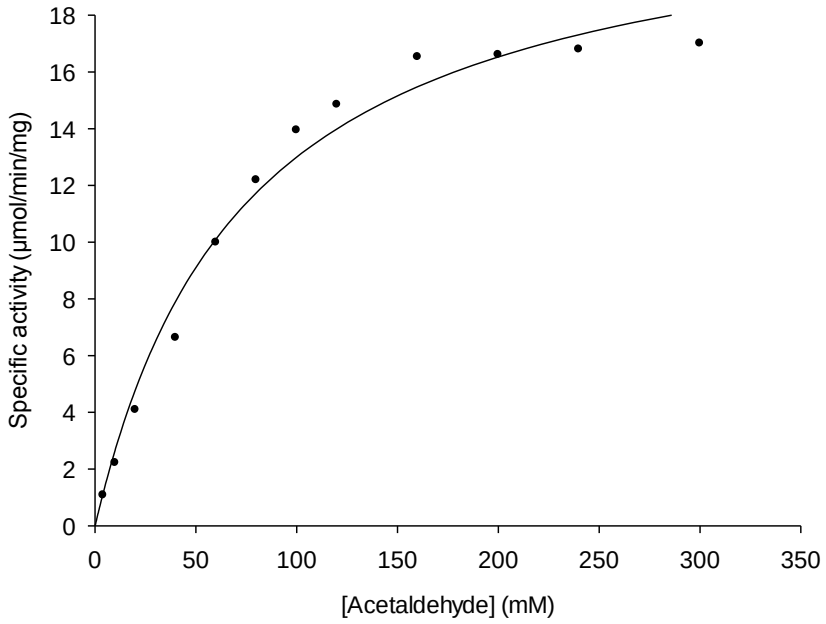


Figure 4-8: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) versus concentration of acetaldehyde (mM) at a fixed concentration of NADH (0.22 mM) in a TM400-Fragment 11 cell extract.

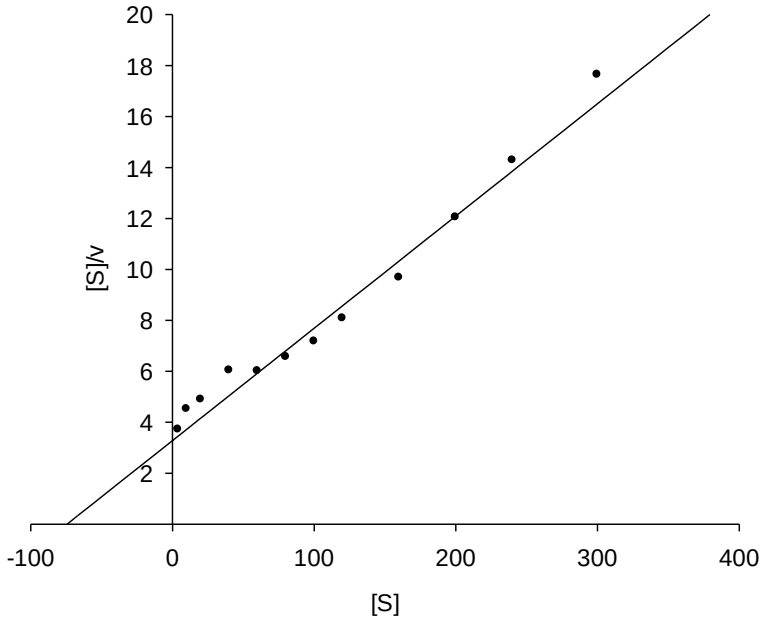


Figure 4-9: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of acetaldehyde (mM) in a TM400-Fragment 11 cell extract.

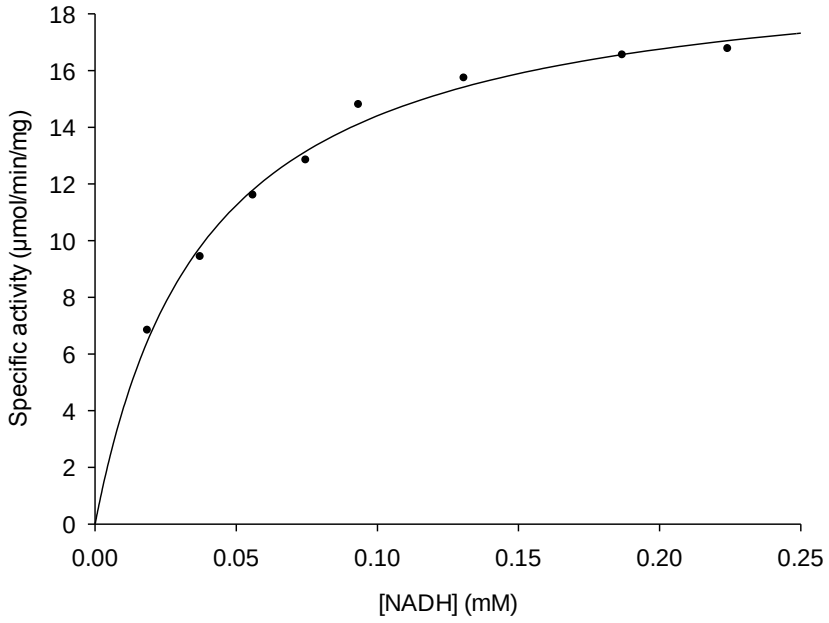


Figure 4-10: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) versus concentration of NADH (mM) at a fixed concentration of acetaldehyde (240 mM) in a TM400-Fragment 11 cell extract.

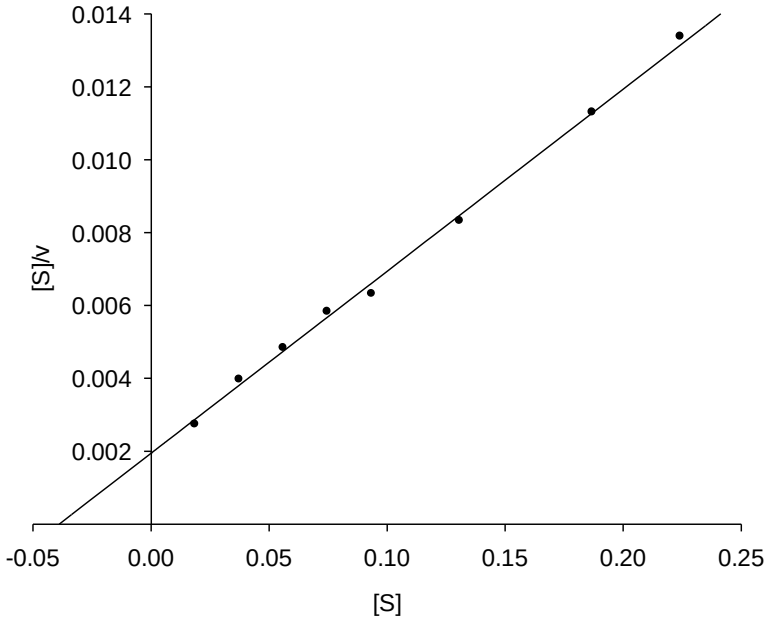


Figure 4-11: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of NADH (mM) in a TM400-Fragment 11 cell extract.

The observed K_m of the protein fragment is similar to that obtained for the ADH domain of the semi-purified ADHE protein which was 79.5 mM (Chapter 3). This observation suggests the absence of the aldDH domain does not significantly affect the activity of the ADH domain of the enzyme with respect to acetaldehyde.

4.3.4 Protein fragment sizing experiments

As described in Section 3.5, the ADHE protein appears to form larger multimeric protein complexes. To determine if the aldDH domain is required for this assembly, the three functional ADH fragments were partially-purified using ion-exchange chromatography prior to sizing by gel filtration (method described in Section 4.2.4).

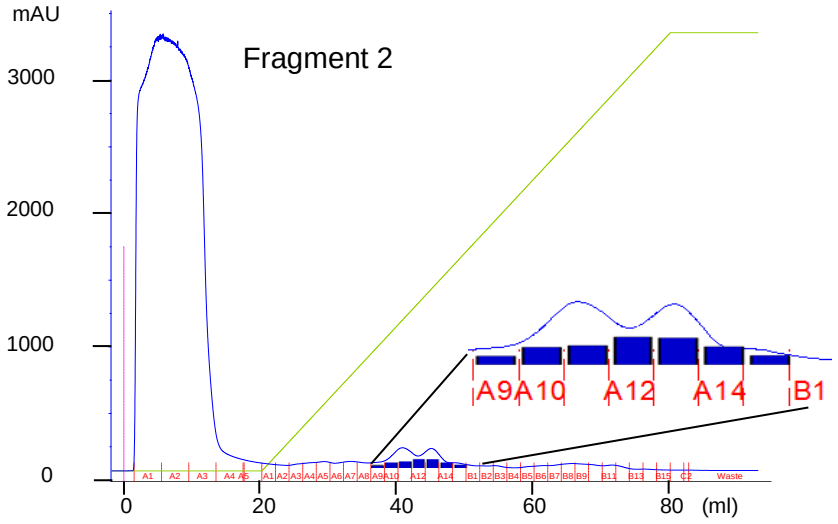


Figure 4-12: Chromatograph of a TM400-Fragment 2 cell extract purification by cation exchange. Blue line = A_{280nm} (mAU), Green line = salt gradient 0-1 M. Peak of activity measurements (abs/min) for ADH is overlaid onto corresponding fractions.

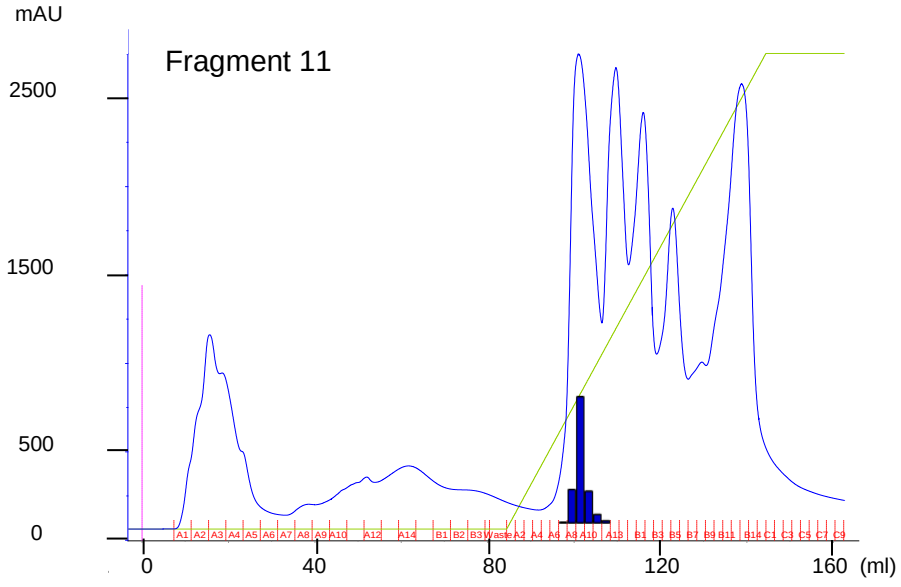


Figure 4-13: Chromatograph of a TM400-ADH Fragment 11 cell extract purification by anion exchange. Blue line = A_{280nm} (mAU), Green line = salt gradient 0-1 M. Peak of activity measurements (abs/min) for ADH is overlaid onto corresponding fractions.

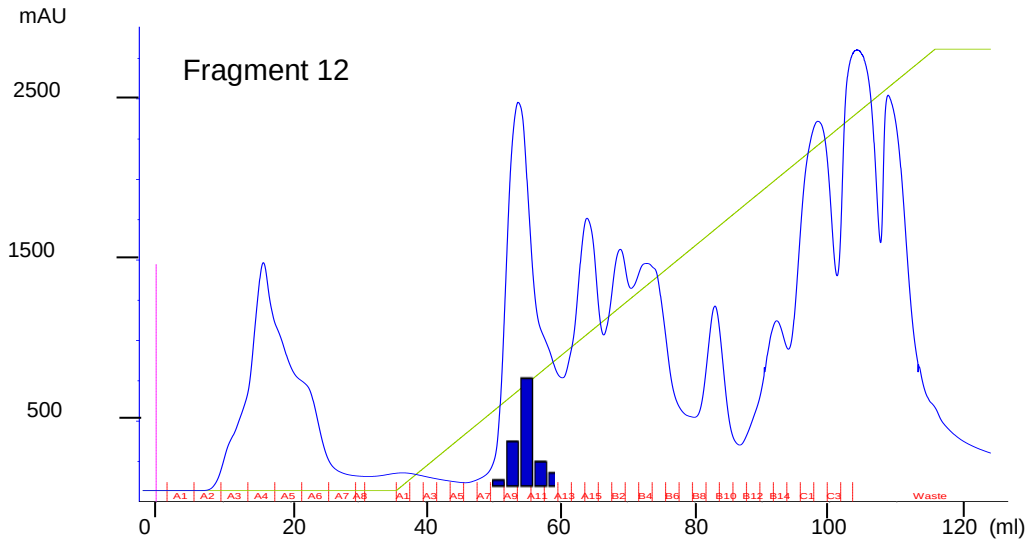


Figure 4-14: Chromatograph of a TM400-ADH Fragment 12 cell extract purification by anion exchange. Blue line = A_{280nm} (mAU), Green line = salt gradient 0-1 M. Peak of activity measurements (abs/min) for ADH is overlaid onto corresponding fractions.

To estimate the size of the protein fragments, the pooled peak fractions from ion-exchange purifications were concentrated and run individually on the gel filtration column. Standard proteins were used to provide an estimate of the sizes of the protein species observed by gel filtration. The estimates of size for the major peaks of activity are shown in Table 4-5.

Protein peak identity		**$V_{elution}$ (ml)**	**$M_{r\ Observed}$**	**$M_{r\ Subunit}$**	**$M_{r\ Observed}$/ $M_{r\ Subunit}$**
Frag 2	peak 1	8.23*	>1,300,000	50,679	>26.0
	peak 2	12.93	171,000		3.4
	peak 3	15.13	66,000		1.3
Frag 11	peak 1	14.58	84,000	46,315	1.8
	peak 2	15.50	56,000		1.2
Frag 12	peak 1	14.90	73,000	67,244	1.1
	peak 2	16.26	41,000		0.6

Table 4-5: Summary of the ADH fragment sizing by gel filtration. * This volume corresponds to elution in the void volume of the column.

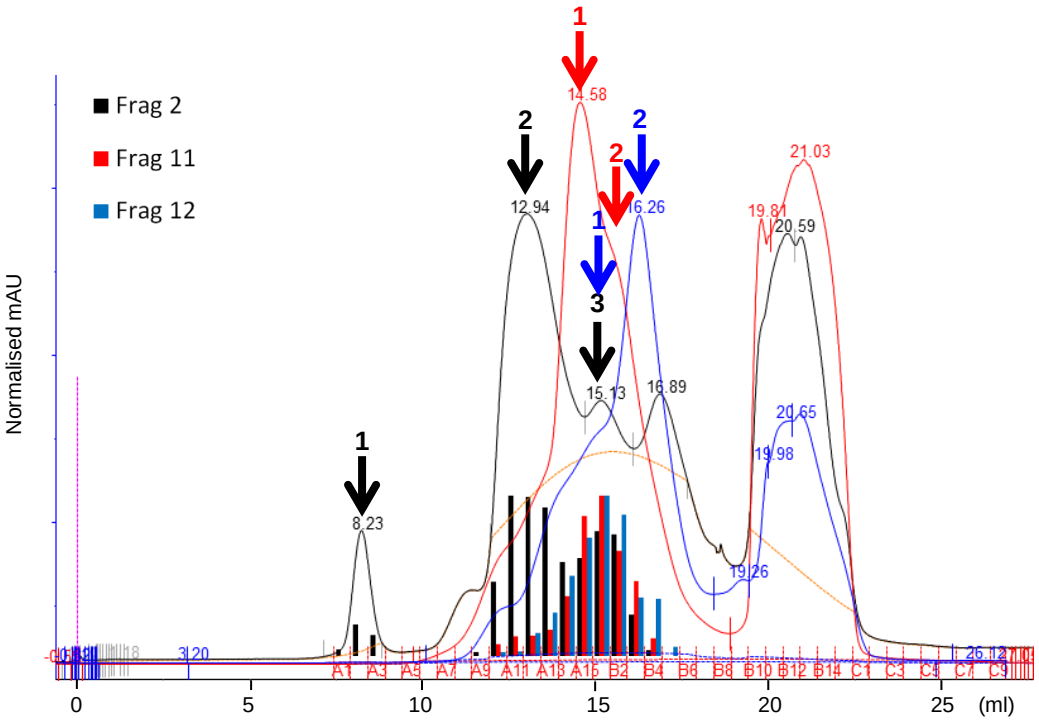


Figure 4-15: Summary chromatograph of TM400-fragment gel filtration runs. Chromatograph lines correspond to normalised A_{280nm} readings (mAU). Peak of ADH activity measurements (% activity of maximum) are overlaid onto corresponding fractions (black = Fragment 2, red = Fragment 11, blue = Fragment 12). Peaks used for size estimation are indicated with the appropriately coloured arrow.

A peak in the void volume of the column, corresponding to the large protein multimers observed for ADHE, was only observed for Fragment 2. This species was relatively low in abundance compared to smaller assemblies (trimers and monomers) within the sample. The absence of the large multimers in Fragment 12 may be due to the large portion of the aldDH domain (potentially disordered) present at the N-terminus interfering with assembly. It would appear that the whole ADHE protein is required to fully form the large multimeric assemblies observed in Section 3.5. Fragment 11, the smallest functional ADH protein, appeared to form dimers and monomers.

4.4 *E. coli* expression host results

To ensure that the lack of detectable aldDH activity was not caused by poor expression using the *Geobacillus* expression system, two of the aldDH fragments (1 & 3) were selected for expression in *E. coli*. Fragments 2 and 11 (ADH) were also selected for recombinant expression to produce greater yields of protein for characterisation and structural investigation. The recombinant expression scheme selected for the proteins incorporated a Histidine tag at the N-terminus to facilitate purification of the proteins produced.

4.4.1 Fragment cloning (pET28a)

PCR amplified DNA fragments were sized (Figure 4-16), and bands corresponding to the expected fragments were excised from agarose gels. Fragments were gel purified, A-tailed and ligated into the pGEM®-T easy vector overnight, prior to ethanol precipitation and transformation into JM109 cells for blue/white screening. Isolated plasmids were purified from overnight cultures of white colonies selected for each fragment. These plasmids were screened using restriction digestion with the appropriate enzymes. Positive results were sent for sequencing.

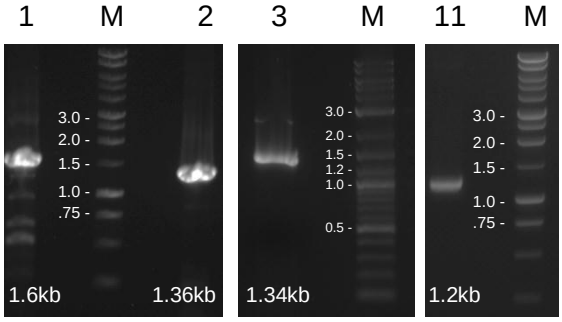


Figure 4-16: Agarose gel electrophoresis of PCR products for His-tagged fragments. Lanes 1-11 = Fragments 1-11, M = DNA markers sizes given in kb.

Once the fragments were confirmed to be successfully cloned into pGEM®-T easy, the DNA fragments were excised using the appropriate restriction enzymes, prior to ligation into the previously-digested and SAP-treated pET28a vector. Ligations were ethanol precipitated prior to transformation into JM109. Transformants were screened using the Kan resistance marker for *E. coli* present in this vector. Representative colonies for each fragment were picked for overnight cultures prior to plasmid preparation. Initially, the vectors were screened using restriction digestion. Positive results were confirmed by DNA sequencing of the fragments using vector-specific primers.

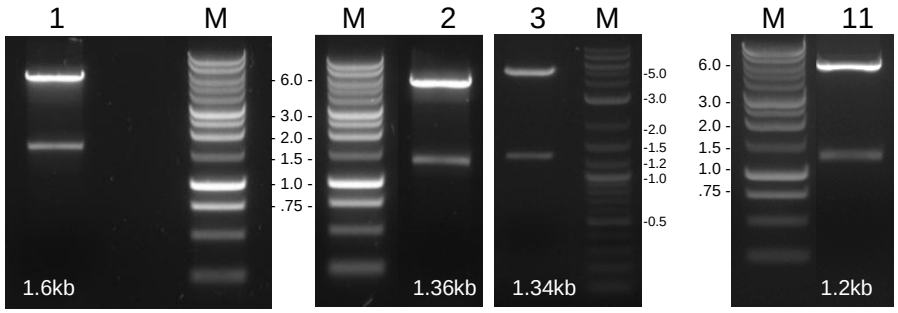


Figure 4-17: Agarose gel electrophoresis of pET28a-Fragments restriction digests. Numbers above = Fragments 1-11, M = DNA markers sizes given in kb. The predicted sizes of the fragments are shown below the respective lane for each fragment. The predicted size of the pET28a vector is 5.4kb.

4.4.2 Recombinant protein expression trials

Purified plasmid constructs were transformed into *E. coli* protein expression strains BL21 (DE3), Rosetta© (DE3) and Arctic Express© (DE3). The time of exposure to IPTG was varied in an effort to optimise protein expression. Suitable levels of protein expression for the ADH fragments were found in BL21 (DE3) cells, following 4.5 h of protein expression conditions. Rosetta© (DE3) cells following 4.5 h of protein expression conditions were used for the aldDH fragments.

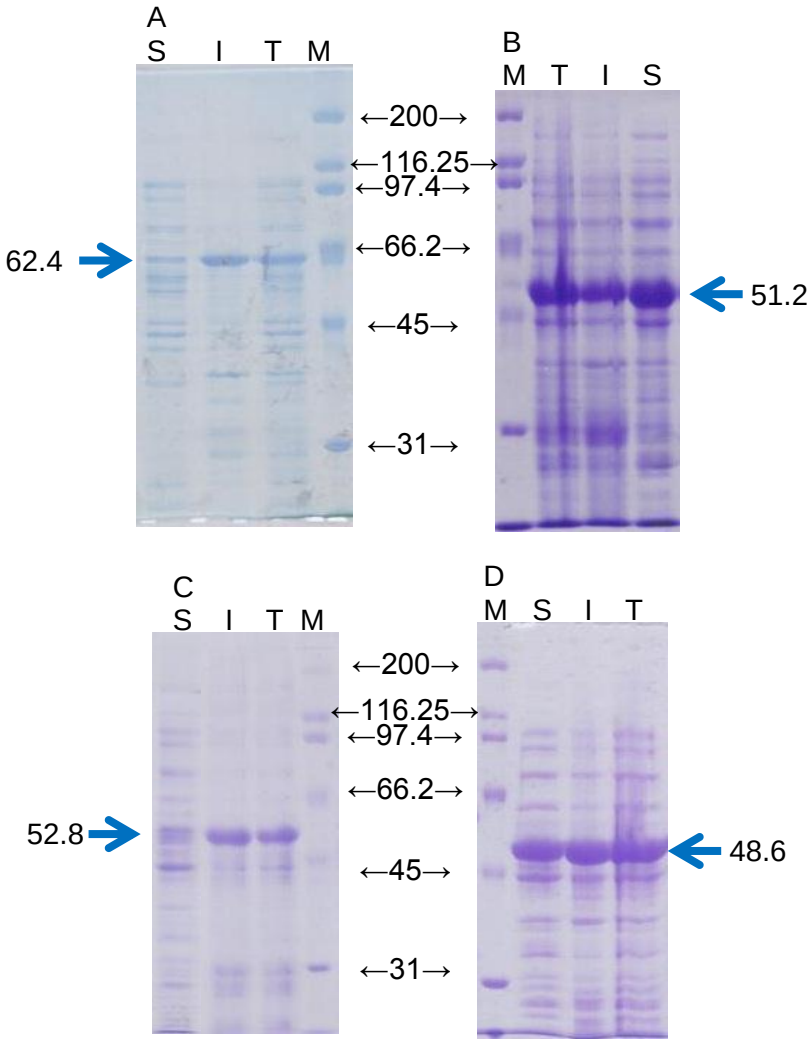


Figure 4-18: SDS-PAGE analysis of Rosetta (aldDH) and BL21 (ADH) fragment protein expression trials. M = markers (M_r/1000), T = total, I = insoluble & S = soluble. A = Fragment 1, B = Fragment 3, C = Fragment 2 and D = Fragment 11. Predicted His-tagged-fragment bands are highlighted with blue arrows, and in each case the expected M_r/1000 is shown.

4.4.3 Protein purification

Extracts of cells expressing the recombinant fragments were prepared as described in Section 2.6. The proteins expressed from the pET28a vector possessed an N-terminal histidine tag; this allowed purification using metal-affinity chromatography (column

charged with Ni^{2+} according to the method outlined in Section 2.9). Conditions were optimised for the respective fragments and proteins were resolved to differing levels of purity.

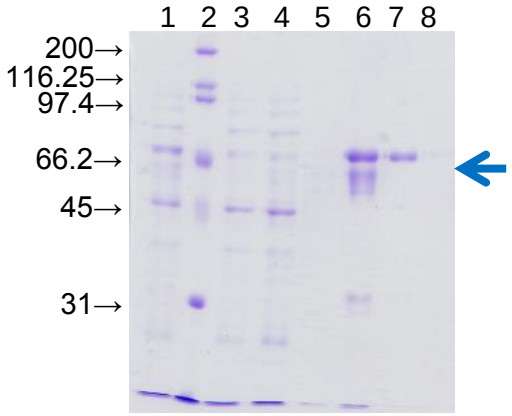


Figure 4-19: SDS-PAGE analysis of Fragment 1 metal affinity chromatography samples. 1 = soluble, 2 = markers (M_r/1000), 3 = flow through, 4 = 0%, 5 = 5%, 6 = 20%, 7 = 40% & 8 = 100%. Predicted His-tagged-Fragment 1 band (M_r/1000 = 62.4) is highlighted with a blue arrow (% values = %HIS-ELUTE buffer diluted in HIS-BIND buffer).

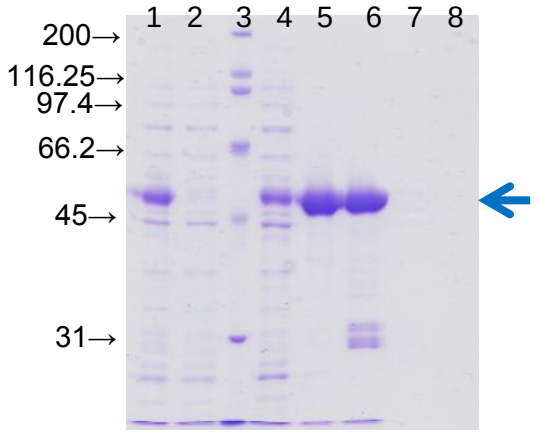


Figure 4-20: SDS-PAGE analysis of Fragment 3 metal affinity chromatography samples. 1 = soluble, 2 = flow through, 3 = markers (M_r/1000), 4 = 0%, 5 = 2.5%, 6 = 30%, 7 = 60% & 8 = 100%. Predicted His-tagged-Fragment 3 band (M_r/1000 = 51.2) is highlighted with a blue arrow (% values = %HIS-ELUTE buffer diluted in HIS-BIND buffer).

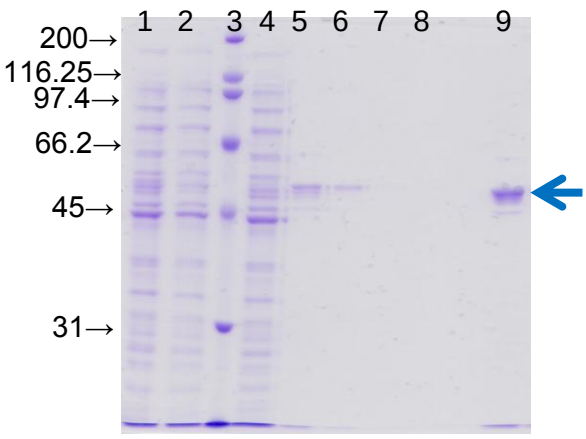


Figure 4-21: SDS-PAGE analysis of Fragment 2 metal affinity chromatography samples. 1 = soluble, 2 = flow through, 3 = markers (M_r/1000), 4 = 0%, 5 = 5%, 6 = 20%, 7 = 40%, 8 = 100% & 9 = Pooled 5+20%. Predicted His-tagged-Fragment 2 band (M_r/1000 = 52.8) is highlighted with a blue arrow (% values = %HIS-ELUTE buffer diluted in HIS-BIND buffer).

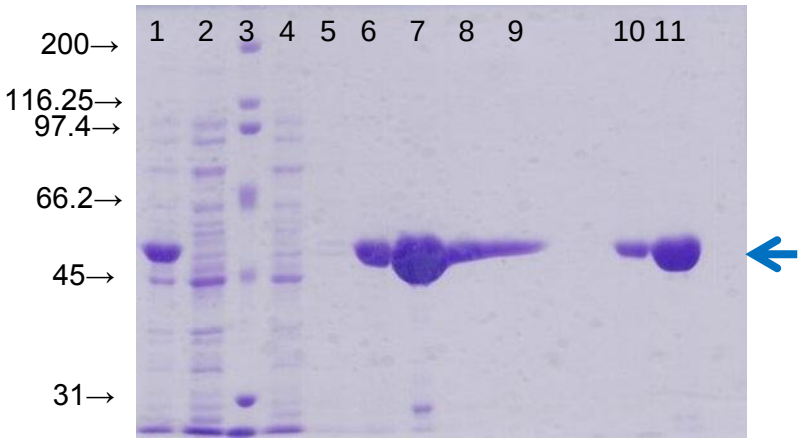


Figure 4-22: SDS-PAGE analysis of Fragment 11 metal affinity chromatography samples. 1 = soluble, 2 = flow through, 3 = markers (M_r/1000), 4 = 0%, 5 = 2.5%, 6 = 10%, 7 = 30%, 8 = 60%, 9 = 100%, 10 = 10% diluted & 11 = 30% diluted. Predicted His-tagged-Fragment 11 band (M_r/1000 = 48.6) is highlighted with a blue arrow (% values = %HIS-ELUTE buffer diluted in HIS-BIND buffer).

Fractions containing the protein fragments of interest (assessed by SDS-PAGE and by enzyme assays where appropriate) were pooled and dialysed into the buffers used for storage overnight at 4°C (Fragments 1, 2 and 3 : 0.2 M phosphate buffer pH 6.0, 0.1 mM zinc acetate; Fragment 11: 50 mM EPPs pH 8.0, 0.1 mM zinc acetate).

4.4.4 AldDH fragments enzyme assays

No aldDH activity was detected for any of the protein fragments during the metal affinity purification or following dialysis. Experiments focussed on the well-expressed Fragment 3 that corresponds to the active aldDH fragment described by Chen et al. (2004) in *E. histolytica*. Incubation with DTT (1 mM) did not have any effect on the

activity of the protein and no activity was detected in assays in the reverse direction (acetaldehyde to acetyl-CoA). The N-terminal tag was removed using the thrombin cleavage site present in this expression system, in case it was interfering with the activity of the protein.

The tag was cleaved using a thrombin cleavage kit according to the manufacturer's instructions, (Novagen) using 0.04 U of thrombin to 0.2 mg of purified Fragment 3 for 6 h at 20.5°C. The protein was run on a metal affinity column and eluted in the flow-through (as confirmed by SDS-PAGE). No activity was detected following the cleavage of the tag in the pre or post metal affinity column samples.

Co-assaying of the ADH Fragment 2 with Fragment 1 and 3 in a 1:1 ratio was also attempted, to see if the presence of the ADH domain was needed for aldDH activity to be restored. AldDH activity was not detected in these assays.

4.4.5 ADH fragments enzyme assays

ADH activity was detected for both recombinantly expressed protein fragments following dialysis into storage buffers. Kinetic characterisation of the two fragments was carried out.

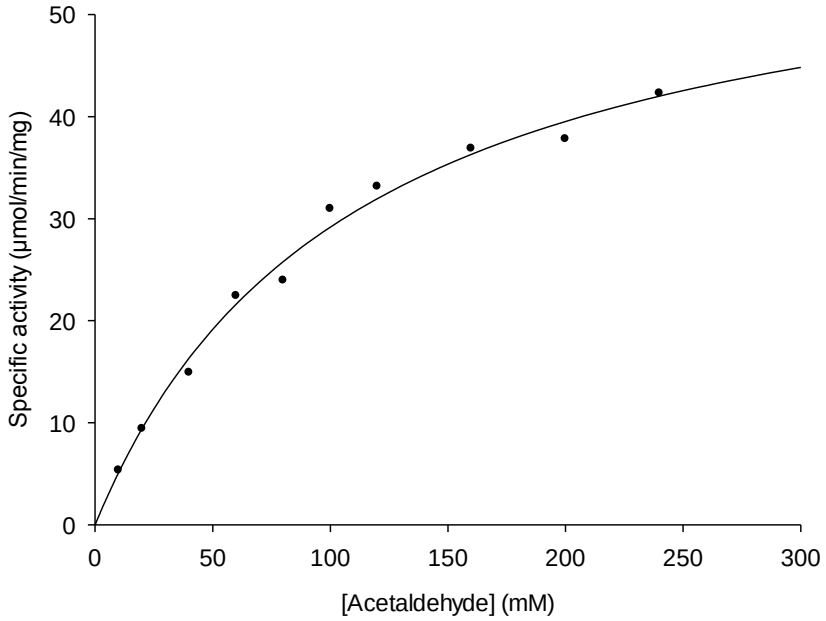


Figure 4-23: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) versus concentration of acetaldehyde (mM) at a fixed concentration of NADH (0.22 mM) for partially-purified recombinant Fragment 2.

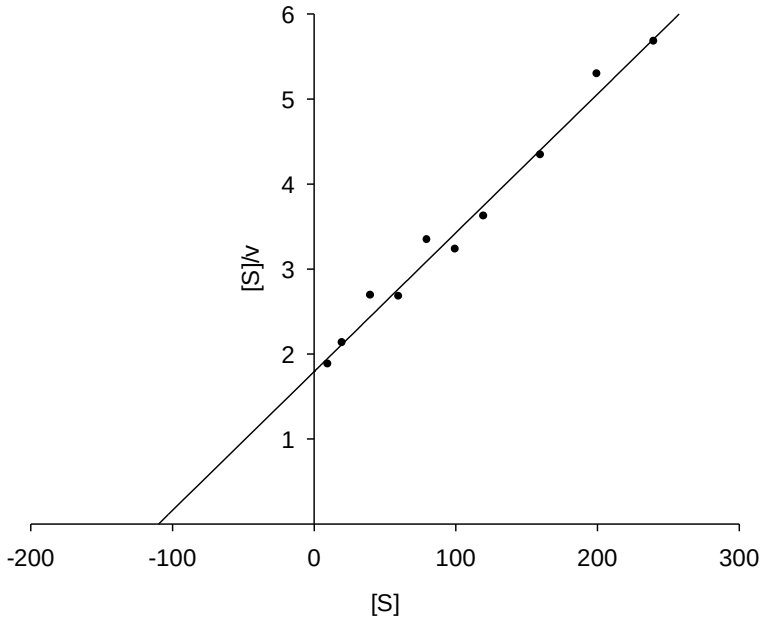


Figure 4-24: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of acetaldehyde (mM) for partially-purified recombinant Fragment 2.

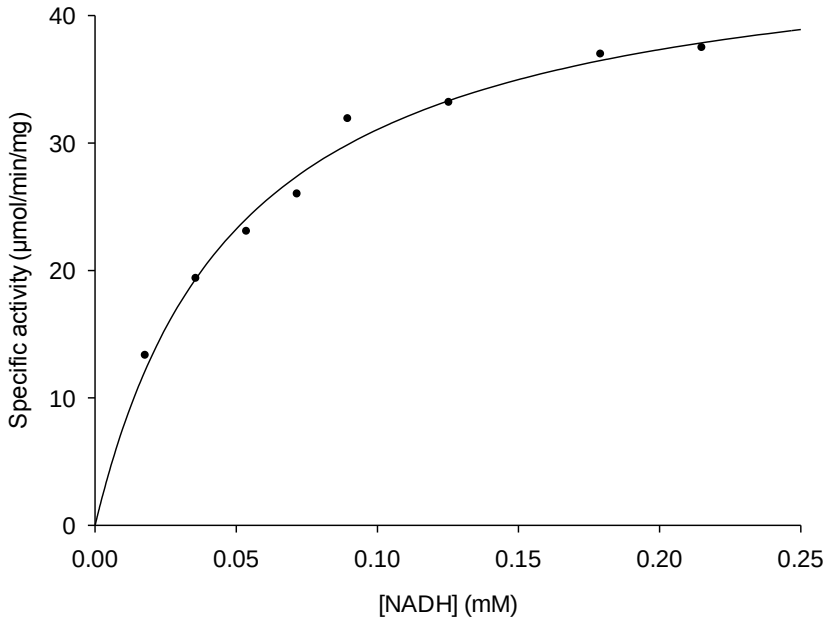


Figure 4-25: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) versus concentration of NADH (mM) at a fixed concentration of acetaldehyde (200 mM) for partially-purified recombinant Fragment 2.

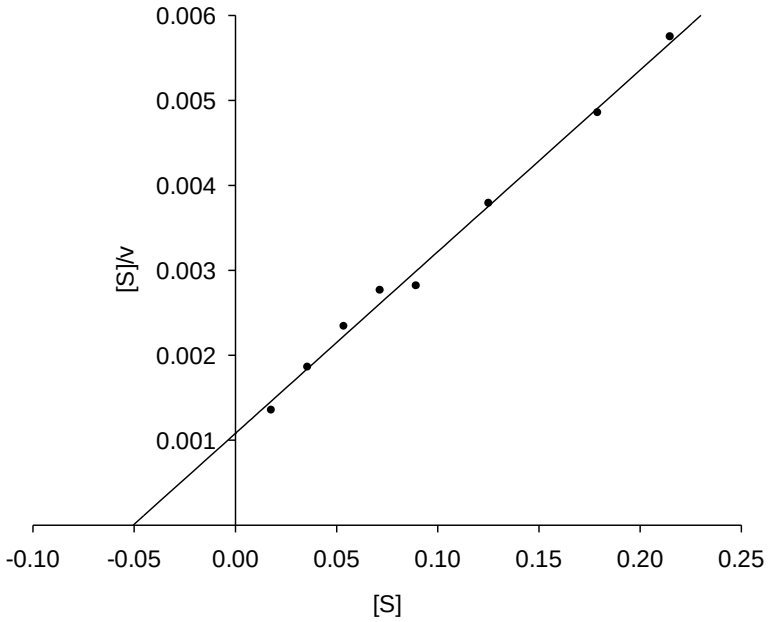


Figure 4-26: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of NADH (mM) for partially-purified recombinant Fragment 2.

Activity	Substrate	Kinetic parameter	Parameter value		Units
ADH	Acetaldehyde	K_m	108.0	+/- (3.8)	mM
		$V_{max\ (adjusted)}$	74.6	+/- (4.8)	U mg^{-1}
ADH	NADH	K_m	0.048	+/- (0.003)	mM
		$V_{max\ (adjusted)}$	71.0	+/- (2.8)	U mg^{-1}

Table 4-6: Kinetic parameters for purified recombinant Fragment 2 protein.

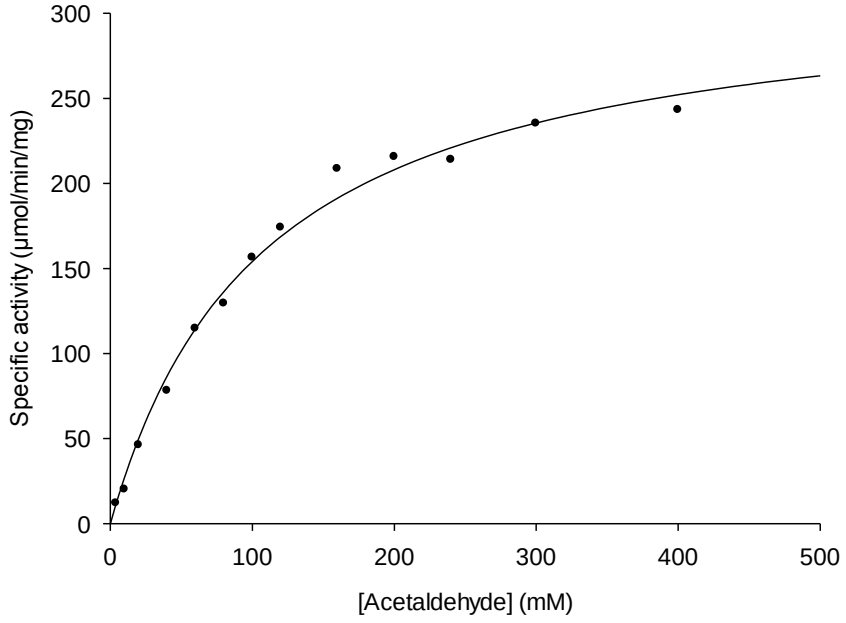


Figure 4-27: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) versus concentration of acetaldehyde (mM) at a fixed concentration of NADH (0.21 mM) for purified recombinant Fragment 11.

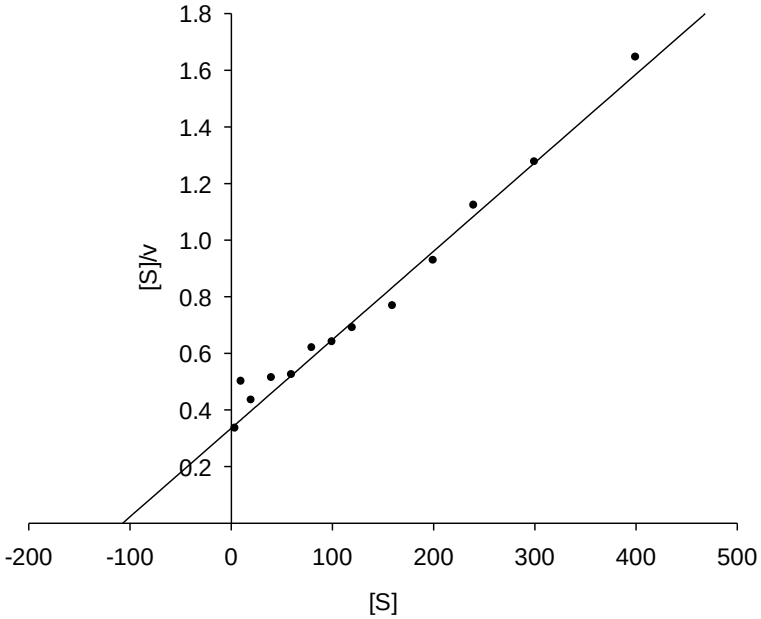


Figure 4-28: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of acetaldehyde (mM) for purified recombinant Fragment 11.

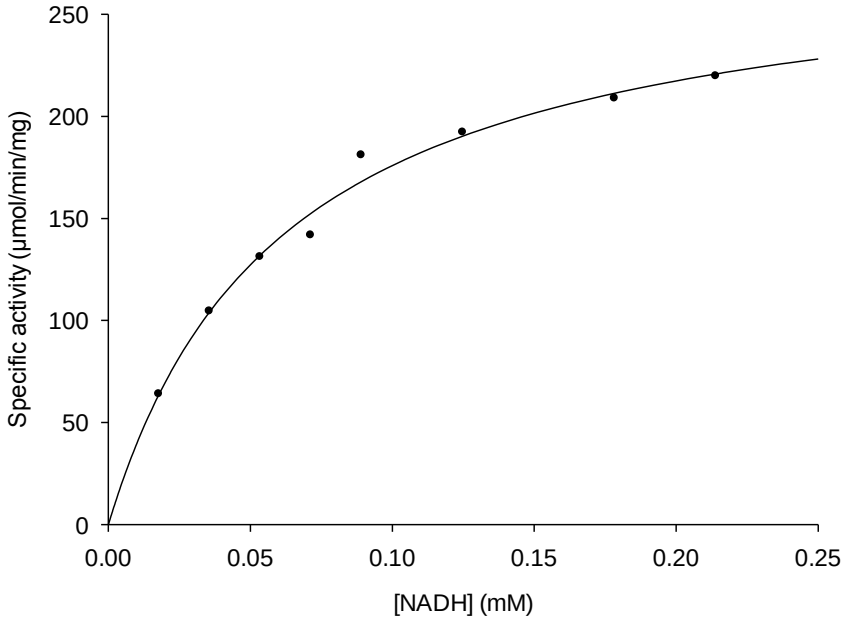


Figure 4-29: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) versus concentration of NADH (mM) at a fixed concentration of acetaldehyde (240 mM) for partially-purified recombinant Fragment 11.

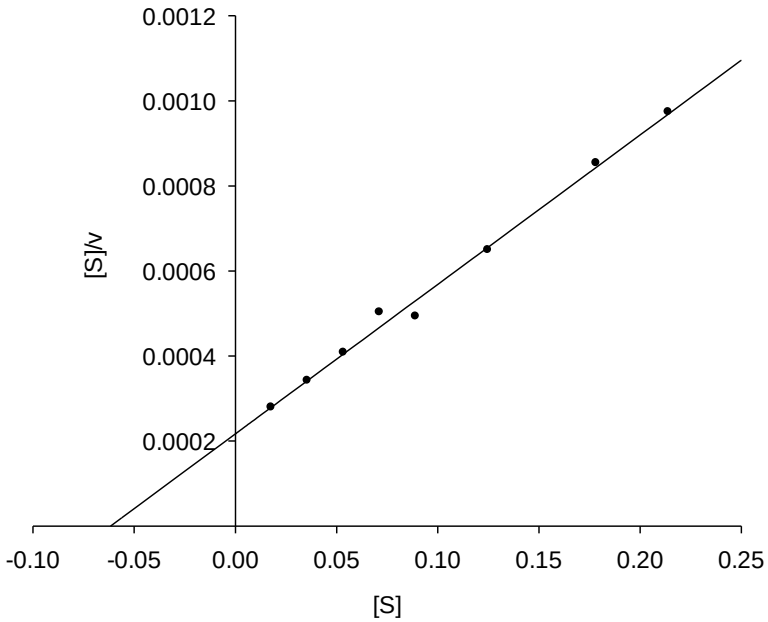


Figure 4-30: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of NADH (mM) for purified recombinant Fragment 11.

Activity	Substrate	Kinetic parameter	Parameter value		Units
ADH	Acetaldehyde	K_m	121.0	+/- (4.5)	mM
		$V_{max\ (adjusted)}$	430.8	+/- (9.7)	U mg^{-1}
ADH	NADH	K_m	0.062	+/- (0.001)	mM
		$V_{max\ (adjusted)}$	427.2	+/- (16.1)	U mg^{-1}

Table 4-7: Kinetic parameter summary data for purified recombinant Fragment 11 protein.

Activity	Substrate	$V_{max\ (adjusted)}$ (U mg^{-1})	K_m (mM)	V_{max}/K_m ((U mg^{-1})/mM)
Fragment 2				
ADH	Acetaldehyde	74.6	108.0	0.7
	NADH	71.0	0.048	1479
Fragment 11				
ADH	Acetaldehyde	430.8	121.0	3.6
	NADH	427.2	0.062	6890

Table 4-8: Summary of catalytic efficiency calculations for recombinant ADH fragments.

Due to the limited purity of the Fragment 2 protein, V_{max}/K_m was used as an indication of catalytic efficiency for the two recombinant ADH proteins. As explained in Chapter 3, comparison of V_{max}/K_m values should be treated with some caution due to limitations surrounding this parameter. SDS-PAGE gels of the Fragment 2 protein showed the

sample to be approximately 80% pure compared to the virtually fully purified Fragment 11 protein. Even when the difference in protein purity is taken into consideration, comparison of the kinetic parameters determined for the two ADH protein fragments showed Fragment 11 to be the more catalytically active of the two proteins. This protein was also expressed at much higher levels by the protein expression strains used.

Further characterisation of the isolated ADH domain of the ADHE protein focussed on Fragment 11.

4.4.6 Fragment 11 optimum temperature assays

Temperature optimum assays were carried out on the purified Fragment 11 protein as described in Section 2.13.5.

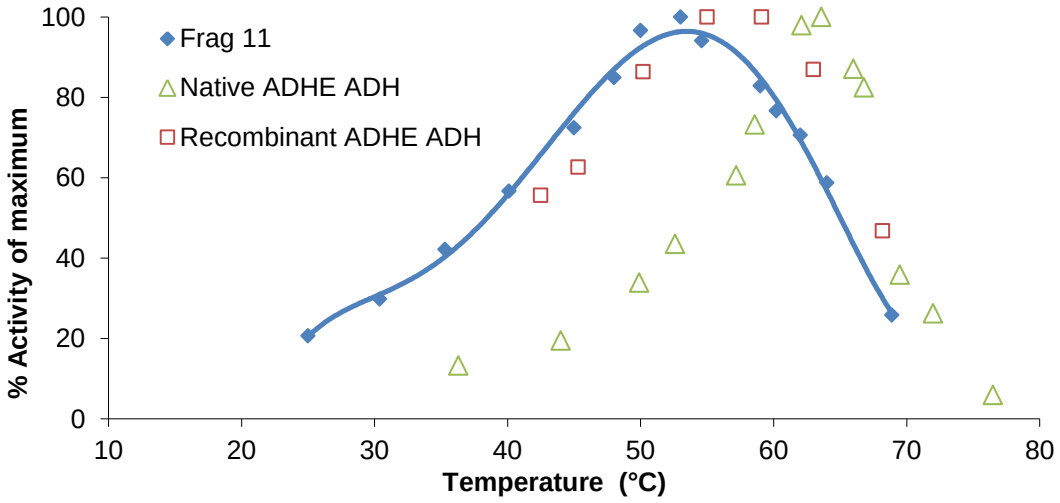


Figure 4-31: Plot of % activity of maximum against temperature (°C) for ADH activity of Fragment 11 compared to native ADHE ADH domain (Chapter 3).

The optimum temperature of the recombinant Fragment 11 protein was 53°C. This is lower than that observed for the native ADHE protein ADH domain (64°C), and the recombinant ADHE protein (57°C).

4.4.7 Fragment 11 thermo-stability assays

Thermo-stability assays were carried out on the purified Fragment 11 protein as described in Section 2.13.6. The protein was split into 100 µl aliquots for incubation at the desired temperatures.

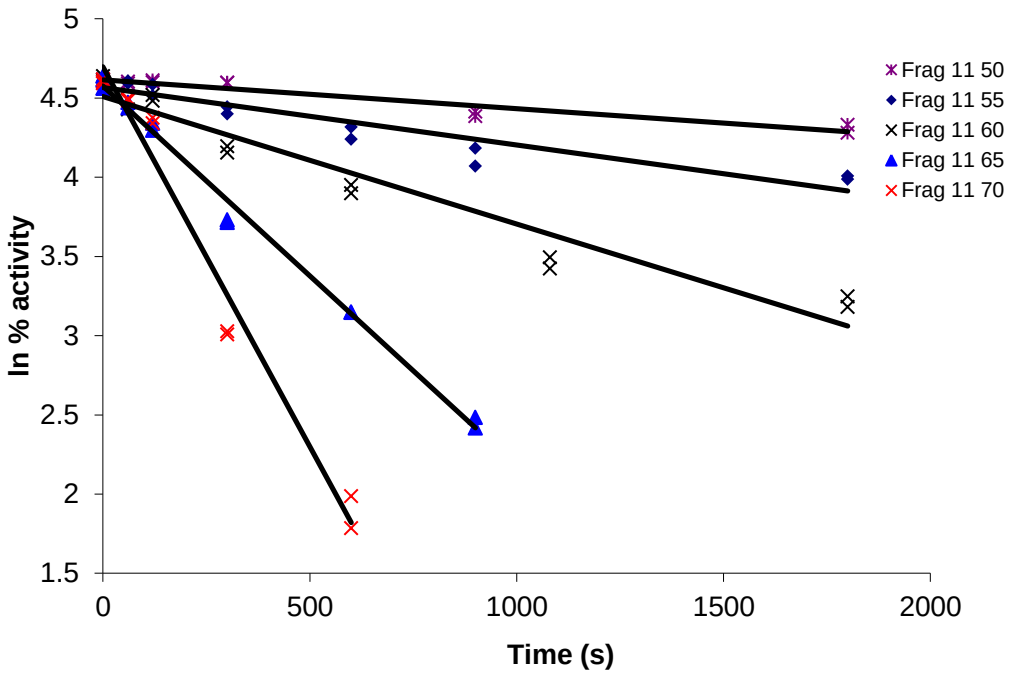


Figure 4-32: Plot of ln % activity of maximum versus time (s) at a defined temperature (°C) for ADH activity of Fragment 11.

The half-life of Fragment 11 at 60°C was approximately 10 min. At 55°C the half-life was in excess of 30 min. Due to the relative impurity of the native ADHE protein compared to purified Fragment 11, the thermo-stabilities could not be compared as other proteins present in the sample may influence protein thermo-stability.

4.4.8 Estimation of the size of Fragment 11

Gel filtration was carried out on purified recombinant Fragment 11 (0.6 ml of 11.9 mg/ml were loaded) in 50 mM EPPS pH 8.0, 5 mM reduced glutathione, 5 mM EDTA and 10% (v/v) glycerol. Assays of ADH activity were used to monitor the presence of Fragment 11. The retention time of the peak of ADH activity was compared to standard proteins to provide an estimate of the M_r.

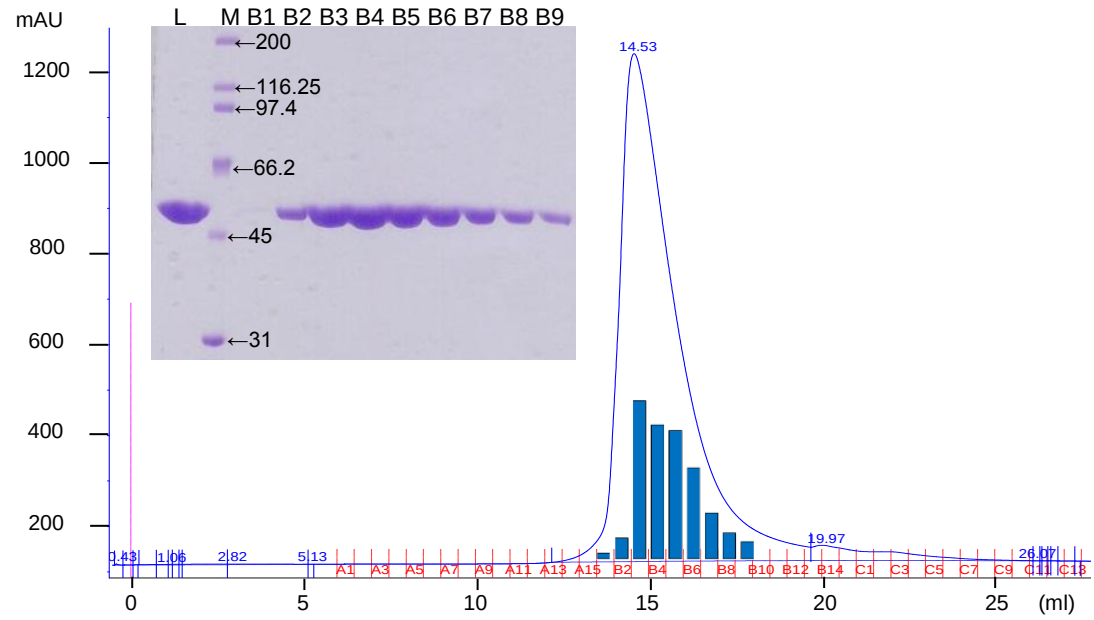


Figure 4-33: Gel filtration of recombinant Fragment 11. Chromatograph line corresponds to A_{280nm} (mAU). Peak of ADH activity measurements (µmol/min/fraction) are overlaid onto corresponding fractions. SDS-PAGE analysis of peak fractions is shown (L = load, M = markers (M_r/1000), B1-B9 = protein fractions from gel filtration). The expected M_r/1000 of Fragment 11 is 48.6.

Protein peak identity		$V_{elution}$ (ml)	$M_{r\ Observed}$	$M_{r\ Subunit}$	$M_{r\ Observed}$/ $M_{r\ Subunit}$
Frag 11	peak 1	14.53	86,000	48,609	1.8

Table 4-9: Fragment 11 sizing by gel filtration.

Fractions B3-7 were pooled and analysed by DLS at 20°C as described in Section 3.2.4, taking into account the viscosity of the buffer used.

			Diam. (nm)	% Volume	Width (nm)
Z-Average (d.nm):	177	Peak 1:	8.49	99.9	1.91
PdI:	0.398	Peak 2:	1020	0.1	248
Intercept:	0.936	Peak 3:	0.00	0.0	0.00

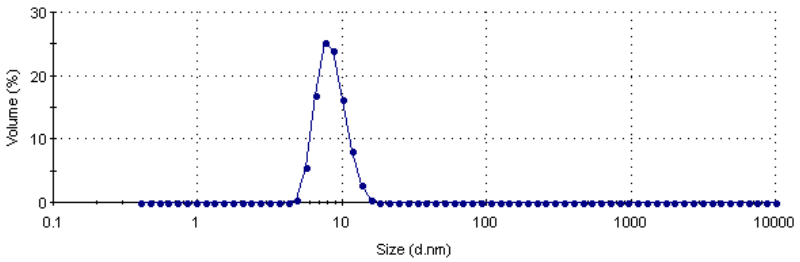


Figure 4-34: Recombinant Fragment 11 DLS analysis result: distribution by volume.

A peak value of 8.49 nm was used for the globular protein estimation. The purified protein sample was shown to be mono-disperse, and the globular protein estimate for this species calculated using the instrument software was 99 kDa (2.03 polypeptides).

Both DLS and gel filtration showed that Fragment 11 appears to be dimeric.

4.4.9 Association experiments

The ADHE protein appears to form large spirosome complexes when produced in the full length form. It is likely that the two domains of the protein interact in this assembly, and this may have implications for protein activity. Fragments 1, 3 and 11 were purified using metal affinity columns prior to dialysis into the running buffer for gel filtration (50 mM EPPS pH 8.0, 5 mM reduced glutathione, 5 mM EDTA and 10% (v/v) glycerol). The proteins were run on the gel filtration column individually, prior to incubating Fragment 11 1:1 with Fragment 1 or 3 for approximately 20 min on ice. These protein mixtures were then re-run on the gel filtration column looking for any shift in the elution volume for the proteins due to association. No shift was observed for either of these protein mixtures suggesting the protein fragments do not associate.

4.4.10 Metal ion stimulation of ADH activity

Experiments to investigate the effect of zinc ions on ADH activity were carried out as in Chapter 3. As shown in Table 4-10, assays of Fragment 11 activity in unfractionated cell extracts (prior to purification) showed ADH activity was stimulated by the addition of zinc to the assay. No ADH activity was detected in a cell extract of a no-vector BL21 control, grown under similar conditions (with and without zinc in the assays), showing the activation to be Fragment 11 specific. As observed for the recombinant ADHE protein in Chapter 3, this modulation of activity was not detected for the post-purification samples.

Sample	Metal ion (0.1 mM)	% ADH activity (of activity with no metal ions)
Cell extract	None	100
	Zn^{2+}	136
Purified Fragment 11	None	100
	Zn^{2+}	102

Table 4-10: Relative effect of metal ions in the assay buffer on the recombinant ADH activity of Fragment 11.

Samples of the purified Fragment 11 protein were subsequently dialysed overnight at 4°C into 50 mM EPPS pH 8.0, +/- 1 mM EDTA; a 75% loss in specific activity was observed for the EDTA-containing sample, which was not restored upon incubation with zinc acetate. Difficulties with protein stability prevented further investigation of metal ion activation.

4.5 Discussion

Fragmentation of the ADHE protein, using both homologous and recombinant expression systems, has successfully resolved the ADH domain of ADHE. The aldDH activity of the ADHE protein could not be resolved independently of the ADH domain.

No active aldDH fragments were obtained using the homologous expression system. Most of the aldDH fragments produced were not well expressed, or were produced insolubly in the *Geobacillus* strains used. It is possible that active aldDH fragments may have been toxic to the expressing cells due to the accumulation of acetaldehyde.

Two of the aldDH fragments were also expressed in *E. coli*. Despite Fragment 3 being well expressed as a soluble protein, neither of these fragments were produced in an active form. This expression was aerobic which may be unfavourable for the aldDH activity, considering the previously observed oxygen sensitivity of the aldDH domain of ADHE. The presence of DTT did not appear to activate the aldDH protein fragments. Possible interference by the N-terminal tag used during protein expression was also tested without successful activation of the aldDH. The hypothesis that the ADH domain may in some way be required to stabilise/activate the aldDH was tested through co-incubation experiments, but again no aldDH activity could be restored.

As a similar protein from *E. histolytica* yielded an active aldDH protein fragment, it was predicted that a similar result may be observed here. Alignment scores between the two ADHE proteins show 57% identity with 74% similarity, so there are significant regions of the proteins that are not homologous. The differences between the two proteins are a possible cause of the differing results. Key residues from the ADH domain could be required for activity/stabilisation of the aldDH domain to retain function. It is likely that such a requirement would be necessary during the protein folding stage and may not be restored through co-incubation of the two domains.

Three of the four ADH domain-containing protein fragments of ADHE produced active ADH proteins. Fragment 11 was the smallest functional ADH protein fragment

produced. This was the most catalytically active of the ADH proteins, as well as being the protein that had the highest expression level, in both *Geobacillus* and *E. coli* strains. The protein appears to be dimeric in nature when produced by both expression systems, as shown by gel filtration and DLS experiments. Larger M_r species were observed for Fragment 2 by gel filtration, which were not observed for Fragments 11 or 12. The additional 40 amino acids between Fragments 2 and 11 may play a role in the multimeric assembly of ADHE. The larger portion of the aldDH domain present in Fragment 12, may interfere with the association of the ADH domains observed for Fragment 2. A majority of the Fragment 2 protein eluted as smaller protein assemblies, i.e. trimers and monomers, so the observed association may be weak.

The K_m observed for Fragment 11 from TM400 (91 mM), is similar to that observed for the partially-purified native ADHE ADH domain (80 mM). Espinosa et al (2001) showed the K_m for acetaldehyde decreased for the ADH domain alone compared to the full-length ADHE protein. This was not observed during these experiments for the homologously-expressed protein.

There was a significant difference in K_m between the TM400 (91.4 mM +/- 2.9) and BL21 (121 mM +/- 4.5) produced Fragment 11 proteins. This may be due to the addition of the N-terminal His-tag, as well as the different temperatures used for protein expression (58°C & 37°C). The difference observed in the K_m for acetaldehyde for the recombinantly expressed Fragments 2 and 11 was relatively small (Fragment 2: 108 mM +/- 3.8 & Fragment 11: 121 mM +/- 4.5). The recombinantly-produced protein fragments both have a significantly lower K_m than the full-length recombinant protein produced in Arctic Express® (201 mM +/- 5.9). The removal of the potentially mis-folded aldDH domain from the ADHE would appear to have a significant effect on the kinetics of the enzyme.

The T_{opt} for the recombinant Fragment 11 (53°C) was lower than the growth temperature of TM242 (60°C). The protein also had a limited thermo-stability around this temperature with a half-life at 60°C of approximately 10 min. The truncation of the protein to resolve the ADH domain alone may have made the protein less stable. This could be due to an increased flexibility due to the lack of stabilisation by the aldDH domain of the protein. Although no association between the inactive aldDH fragments and Fragment 11 was observed, a physical link between the two domains may be

required for such interactions to form. Structural studies looking at the interactions between the two domains of ADHE would allow this theory to be tested further.

The recombinantly produced Fragment 11 protein could be purified to homogeneity at high yields, making it suitable for structural studies using X-ray crystallography. This was used in an effort to further investigate this domain of the ADHE protein, in terms of protein interactions and key amino-acid residues involved in catalysis. This work is discussed in Chapter 5.

Experiments reported in Chapter 3 and in this chapter suggest that the ADH activity of ADHE, although present without additional metal ions in the assay, can be increased upon the addition of divalent metal ions such as zinc. Stimulation was observed in cell extracts of TM242, in extracts of ArcX cells expressing ADHE and in extracts of BL21 cells expressing Fragment 11; however, modulation of activity was not observed for the recombinant proteins post metal-affinity purification. It would appear that the protein is capable of scavenging divalent metal ions from the column, thus limiting the stimulation of activity during subsequent assays. A metal ion has been identified in the active site of the Fragment 11 protein and is reported in the subsequent chapter describing its structural resolution, implying that catalysis may be directly influenced by the presence of a divalent metal ion. The decrease in activity following metal ion stripping with EDTA also suggests that the presence of metal ions may have a direct effect on catalysis.

In conclusion, experiments carried out during this section of the project have successfully resolved the ADH domain of the ADHE protein, and shown it to be active independently of the aldDH. The minimal functional unit appeared to be 411 amino acids in length and form dimers. ADH fragments appeared to retain the relatively high K_m for acetaldehyde observed with the ADHE protein. The aldDH domain remains unresolved despite several different sized fragments being tested and attempts at reactivation being carried out.

5 CRYSTALLISATION OF THE MINIMAL FUNCTIONAL ADH DOMAIN (FRAGMENT 11)

5.1 Introduction

Although several attempts were made, the ADHE protein from *G. thermoglucosidasius* was not crystallised successfully (see Section 5.3). The presence of a flexible linker between the two domains of the protein, limitations in terms of protein purity, and the presence of the multimeric spirosome structures are possible explanations for the lack of crystallisation screen hits. The ADH domain of ADHE in isolation was therefore investigated.

No published protein structures were available for the ADH domain of an ADHE protein at the time this work was carried out. The smallest functional fragment of ADHE corresponding to the ADH domain of the protein was Fragment 11. This protein was highly soluble and purified readily from recombinant expression strains using the N-terminal His-tag incorporated with the pET28a vector. Structural analysis of the ADH domain of the ADHE protein was carried out in order to elucidate various properties of the domain, such as its potential interactions with the aldDH domain of ADHE, the presence/absence of divalent metal ions and the catalytic/structural properties of any metal ions present.

Also of interest was the recent observation that point mutations in the ADHE protein of *Clostridium thermocellum* can increase the ethanol tolerance of these strains (Brown et al. 2011). Mutations at key residues appeared to reduce ADH activity and switch its cofactor specificity from NADH to NADPH. The structural basis of their work was models generated from similar structures (a 1,3-propanediol dehydrogenase from *Klebsiella pneumoniae* (3BFJ) and an iron containing ADH from *Thermotoga maritima* (1O2D)). Alignments were carried out between the predicted ADH domain of *C. thermocellum*, Fragment 11 and the two proteins used in the paper (Brown et al. 2011); these showed Fragment 11 to have a much higher sequence identity and similarity to the ADH domain of *C. thermocellum* than the models used.

Protein sequence	% Identity	% Similarity	% Gaps
TM242 Fragment 11	50	70	2
Klebsiella pneumoniae 1,3-propanediol dehydrogenase (3BFJ)	32	48	12
Thermotoga maritima Fe-ADH (1O2D)	22	42	16

Table 5-1: Alignment scores for Fragment 11 and two protein used in Brown et al. (2011) for modelling the ADH domain of *C. thermocellum* ADHE (residues 455-873).

To better understand the nature of these mutations and the effect they may have in TM242, a high resolution crystal structure of Fragment 11 was desirable.

Crystallisation of the non-functional aldDH protein Fragment 3 was also attempted during this project; however, these efforts were unsuccessful.

The X-ray crystallography work described here and subsequently in Chapter 7 was carried out under the supervision of Dr Susan Crennell (University of Bath). The author wishes to acknowledge assistance with this work, particularly the final refinement stages of the described structures.

5.2 Materials and Methods (X-ray crystallography)

X-ray crystallography is a widely-used technique within protein biochemistry to determine the three dimensional structure of the protein of interest. To obtain a high resolution crystal structure of a particular protein, an ordered crystal made up of the repeating protein units must be obtained that will adequately diffract X-rays. These diffraction data are used for the determination of the structure of the protein.

For protein crystals to nucleate and grow, the protein and precipitant concentrations, as well as other factors i.e. buffer type, concentration of metal ions, pH, temperature and precipitant type, must all allow the protein to be under supersaturated conditions. If these conditions are not met then no crystal will form, and the protein will either remain soluble or precipitate out of solution (shown in Figure 5-1).

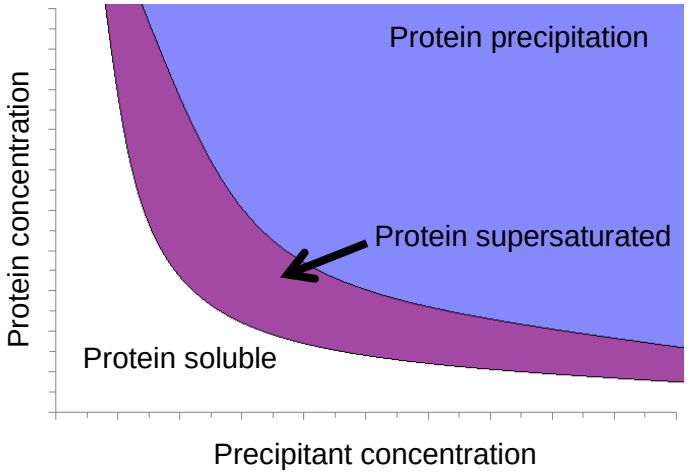


Figure 5-1: Theoretical conditions for crystal nucleation and growth.

The conditions required for crystal growth vary from protein to protein and are influenced by a wide variety of factors such as temperature, surface charge and hydrophobicity.

The method used in this project to achieve supersaturated conditions was vapour diffusion. This technique uses a hanging/sitting drop of protein solution, which is mixed in different ratios with the particular precipitant-containing well solution used for the experiment. The drop is suspended above the un-diluted well solution and left to equilibrate. Water diffuses from the drop containing the protein to the undiluted solution below, gradually changing the conditions experienced by the protein in the drop. If successful, this allows supersaturation conditions for the protein of interest to be met and nucleation followed by crystal growth to follow. A wide variety of conditions are commonly screened for initial crystal "hits", which are then optimised in an attempt to obtain the best conditions for well-ordered crystal growth.

Unless otherwise stated, crystallisation-specific chemicals and consumables used were supplied by Molecular Dimensions (UK).

5.2.1 Pre-crystallisation tests

To ensure the concentration and purity of protein used for crystallisation screens was within an appropriate range, protein samples were subjected to the pre-crystallisation test (PCT) according to the manufacturer's instructions (Hampton Research, USA).

5.2.2 Crystallisation screens

Many high-throughput crystallisation screens are commercially available, ranging from sparse matrix screens to more complex screens with a variety of different additives hypothesised to aid crystal formation. The preliminary screens performed during this project were carried out using a Phoenix crystallisation robot (Art Robbins, UK) in the 96-well format. Plates used for this screening employed the sitting-drop method. Three drops were normally used per well solution for the screens tested. The proportion of protein to well solution was varied between the drops to maximise the potential for reaching supersaturation conditions. The wells were sealed with plastic film prior to incubation at 16°C.

5.2.3 Optimisation of crystallisation conditions

Hits from the initial high-throughput screening conditions were scaled up to larger drop sizes using the hanging drop method in 24-well XRL crystallisation plates, with plastic coverslips sealed with vacuum grease. Drop sizes of 3 µl of a variety of ratios of protein to well solution were suspended above 500 µl of well solution in the plates. Optimisation was generally based around gradients of pH, precipitant concentration and additive concentration depending on the conditions of the best initial hits. Trays were stored at 16°C during the timescale of the experiments.

5.2.4 Data collection

Unless otherwise stated, data were collected at the Diamond light source (Oxford, UK). Crystals were transferred using a cryo-loop from the drop in which they had grown, into a fresh drop containing the well solution with the appropriate concentration of cryo-protectant added. Crystals were then mounted onto the collection apparatus within a stream of nitrogen gas to freeze the crystals. Data collection is described in the appropriate results section of this thesis and was processed using the HKL2000/0.98 or Xia2 software packages.

5.2.5 Molecular replacement and model refinement

BALBES (Long et al. 2008) was used for molecular replacement identifying the best structures to use from the PDB. The model was automatically re-built and refined using the ARPwARP programme (Langer et al. 2008). The COOT software programme (Emsley et al. 2010) was used for manual model building, refinement was carried out using the Refmac5 programme (CCP4i (Potterton et al. 2003)) and model evaluation was carried out using MolProbity (Chen et al. 2010). The cycle of model refinement and evaluation is shown in Figure 5-2.

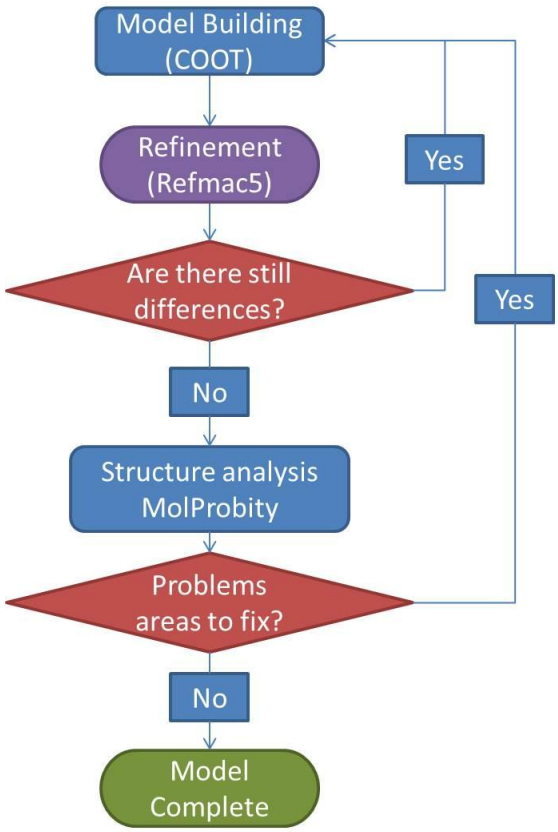


Figure 5-2: Flowchart of model refinement using collected data.

5.2.6 Metal ion analysis

A Perkin Elmer AAnalyst 100 Atomic Absorption spectrometer was used to investigate the presence of zinc and iron within Fragment 11 protein samples. The instrument was calibrated as required with zinc standard solutions at 0.2, 0.4, 1 and 2 parts per million or iron standard solutions 2.5, 5 and 7.5 parts per million. The appropriate dialysis buffer was analysed as a protein-free blank prior to sample analysis. Analysis was carried out by Mr Alan Carver (University of Bath).

Protein samples were previously dialysed overnight at 4°C into 50 mM EPPS pH 8.0, 0.1 mM zinc acetate, prior to dialysis into 50 mM EPPS pH 8.0, with two changes of dialysis buffer overnight at 4°C. Where required, protein samples were diluted in dialysis buffer prior to measurements being taken. The sample was drawn into the instrument and allowed to equilibrate until a steady value for ion concentration was obtained. Concentrations of each protein analysed were determined using the A_{280nm} method described in Section 2.8.

Samples of protein were also analysed by Scanning Electron Microscope-Energy Dispersive X-ray Spectroscopy (SEM-EDS) analysis, using a JSM6480LV scanning

electron microscope (JEOL, UK) fitted with an INCA X-ray analyser (Oxford instruments, UK). In this process, the incident electron beam (15keV) is targeted onto the sample causing collisions with electrons present in the sample. X-rays are released when some of these displaced electrons are replaced with higher energy electrons and excess energy is released as X-rays. Analysis of the emitted X-rays provides an indication of the atoms present within a sample. Protein samples were dried onto graphene-coated sample mounts and the residue was analysed. The instrument was optimised using a copper standard at the same energy settings as the samples. Analysis was carried out by Dr John Mitchels (University of Bath).

5.2.7 Thrombin cleavage of His-Tag

A thrombin cleavage site was incorporated between the His-tag and the start of the Fragment 11 protein sequence when expressed from the pET28a vector. The tag was removed from the purified Fragment 11 protein where required using the following method:

The protein purified as in Section 4.4.3 was dialysed overnight at 4°C into 20 mM Tris-HCl pH 8.4, 2.5 mM $CaCl_2$, 150 mM NaCl and 20 mM imidazole. The protein was digested with 1.2 U mg^{-1} of thrombin for 16 h at 20°C, prior to re-purifying using metal affinity chromatography (replacing the HIS-BIND buffer with the buffer described here and the HIS-ELUTE buffer with 20 mM Tris-HCl pH 8.4, 2.5 mM $CaCl_2$, 150 mM NaCl and 1 M imidazole). Fragment 11, which eluted through washing with 0% HIS-BIND (i.e. did not bind to the column), was assayed to confirm activity and dialysed into the required buffer for subsequent analysis.

5.3 Unsuccessful crystallisation trials

Samples of the partially-purified native ADHE protein (Section 3.3.8.2) in 50 mM EPPS pH 8.0, 0.1 mM zinc acetate, 5 mM GSH and 10% (v/v) glycerol were screened for crystallisation. Structure screen I & II, PGA, JCSG-plus and Proplex screens (Molecular Dimensions, UK) in a 1:1 ratio of sample to well solution (300 nl drop size), at 2 different protein concentrations (4.45, & 2.51 mg/ml), were used.

No crystal hits were obtained for any of the conditions tested. A majority of the conditions resulted in formation of amorphous precipitates.

The purified His-tagged Fragment 3 protein dialysed into 50 mM Tris-HCl pH 8.0, 150 mM NaCl, was screened for crystallisation with Structure screen I & II, and PGA

screens (Molecular Dimensions, UK) in a 1:1 ratio with well solution (300 nl drop size) at 2 different protein concentrations (7.2 & 3.7 mg/ml). The 7.2 mg/ml sample was also screened with the JCSG-plus and Proplex screens (Molecular Dimensions, UK) in a 1:1 ratio with well solution (300 nl drop size). A 7.2 mg/ml sample dialysed into 50 mM Citric acid buffer pH 6.0, 0.1 mM zinc acetate, was also screened with Structure screen I & II, & PGA screens (Molecular Dimensions, UK) in a 1:1 ratio with well solution (300 nl drop size).

No crystal hits were obtained for any of the conditions tested. Granular precipitate was commonly observed as well as amorphous precipitate.

5.4 Fragment 11 crystallisation results

5.4.1 Protein crystallisation

The purified His-tagged Fragment 11 protein dialysed into 50 mM EPPS pH 8.0, 0.1 mM zinc acetate, was screened for crystallisation with Structure screen I & II, PGA, JCSG-plus and heavy and light twin-pack screens (Molecular Dimensions, UK) in a 1:1 ratio with well solution (400 nl drop size) at 3 different concentrations of protein (6.3, 3.3, & 1.8 mg/ml).

The best hit obtained was for the PGA screen well C4 at a protein concentration of 6.3 mg/ml. This condition contained 0.1 M sodium acetate pH 5.0, 0.1 M ammonium sulphate, 0.3 M sodium formate, 10% (v/v) Peg2K MME and 3% (v/v) poly-glutamic acid (PGA-LM).

Optimisation of the conditions employed various gradients of Peg2K MME, PGA-LM, pH and protein concentration to obtain the most suitable crystals for X-ray diffraction. Crystals producing the best diffraction data were grown at a protein concentration of 5.4 mg/ml diluted 1:1 with well solution (0.1 M sodium acetate pH 5.0, 0.1 M ammonium sulphate, 0.3 M sodium formate, 11.5% (v/v) Peg2K MME, and 3-3.5% (v/v) PGA-LM). The crystals produced were flat and diamond shaped in morphology. Crystals took approximately 1 week to appear at 16°C and growth stopped after 2 weeks; representative crystal forms are shown in Figure 5-3. Glycerol (30%, v/v) was required as a cryo-protectant when the crystals were frozen for data collection.

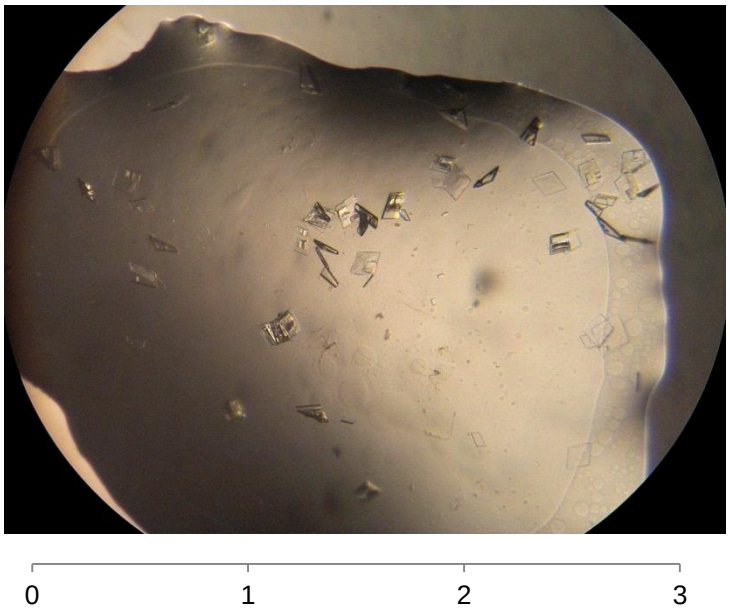


Figure 5-3: Fragment 11 representative crystal forms (scale shown = mm).

5.4.2 Data collection and molecular replacement

Data were collected at the Diamond light source beam-line I03 at a wavelength of 0.97625Å. Data were processed using the Xia2 (Winter 2010) software package. Statistics of data collection are shown in Table 5-2.

Parameter	Value
Unit cell dimensions	a = 73.43 Å b = 96.28 Å c = 58.00 Å α=β=γ = 90.00°
Space group	$P2_12_12$
R_{merge}	0.063 (0.618)
Completeness	99.5 % (99.9)
I/σI	13.4 (2.0)
Multiplicity	4.5 (4.6)
Number of reflections used	66581 (4959)
Number of unique reflections	14837 (1085)
Resolution	40.26-2.49 Å (2.56-2.49)

Table 5-2: Fragment 11 data collection statistics (numbers in brackets for highest resolution bin).

The processed data were submitted to BALBES, which suggested the structure of *E. coli* lactaldehyde reductase (PDB code 1RRM) be used for molecular replacement (99% likely as structure). The protein sequence of the 1RRM structure is 32% identical and 51% similar (8% gaps) to the non-His-tagged Fragment 11 sequence. This

solution was submitted to ARPwARP to improve the fit of the model to the X-ray data collected.

5.4.3 Data refinement

Following creation of the model, the data were subjected to a rigid body refinement in Refmac5. At this point in the refinement process, the R factor was 0.2598 and R_{free} was 0.3184.

The model was subjected to several rounds of refinement and validation, 20 water molecules, a sulphate ion, a Zn^{2+} ion and a glycerol molecule were all added to regions of un-modelled density, and the model refinement process continued until the solution was judged to be complete. The final model excluded one loop region in the structure (discussed later) as there was not sufficient density for the conformation of this loop to be determined. Statistics for the refined final model are shown in Table 5-3.

Parameter	Value
Number of reflections used	14056 (5% test set)
Number of protein atoms	3182
Number of non-solvent atoms	12
Number of solvent molecules	20
Mean B-factors (Å^2)	
Overall	58.06
Protein main chain	54.0
Side chains	60.9
Water molecules	55.3
R	0.1728
R_{free}	0.2624
RMS Bond Length	0.012 Å
RMS Bond Angle	1.557°

Table 5-3: Final refinement statistics for the Fragment 11 crystal structure.

The structure was further evaluated using MolProbity.

Parameter	Value
Ramachandran favoured	96% (382/398)
Ramachandran allowed	99.7% (397/398)
Ramachandran outliers	0.25% (1/398)
MolProbity score	2.45 (78th percentile n=6897, 2.49Å ± 0.25Å)
Clashscore	15.92 (84th percentile n=269, 2.49Å ± 0.25Å)

Table 5-4: MolProbity validation results for Fragment 11.

The only Ramachandran outlier present is Thr_{188}, which is located at the end of the most mobile loop visible within the structure (loop F in Figure 5-13). The density observed for this loop was of lower quality than the rest of the structure. Predictions of residue conformations could be made from the limited density due to the relatively short loop length. It is not surprising that Thr_{188} is an outlier due to this limited density. However, as the residue is not located within the active site, the implications of this residue being an outlier are limited.

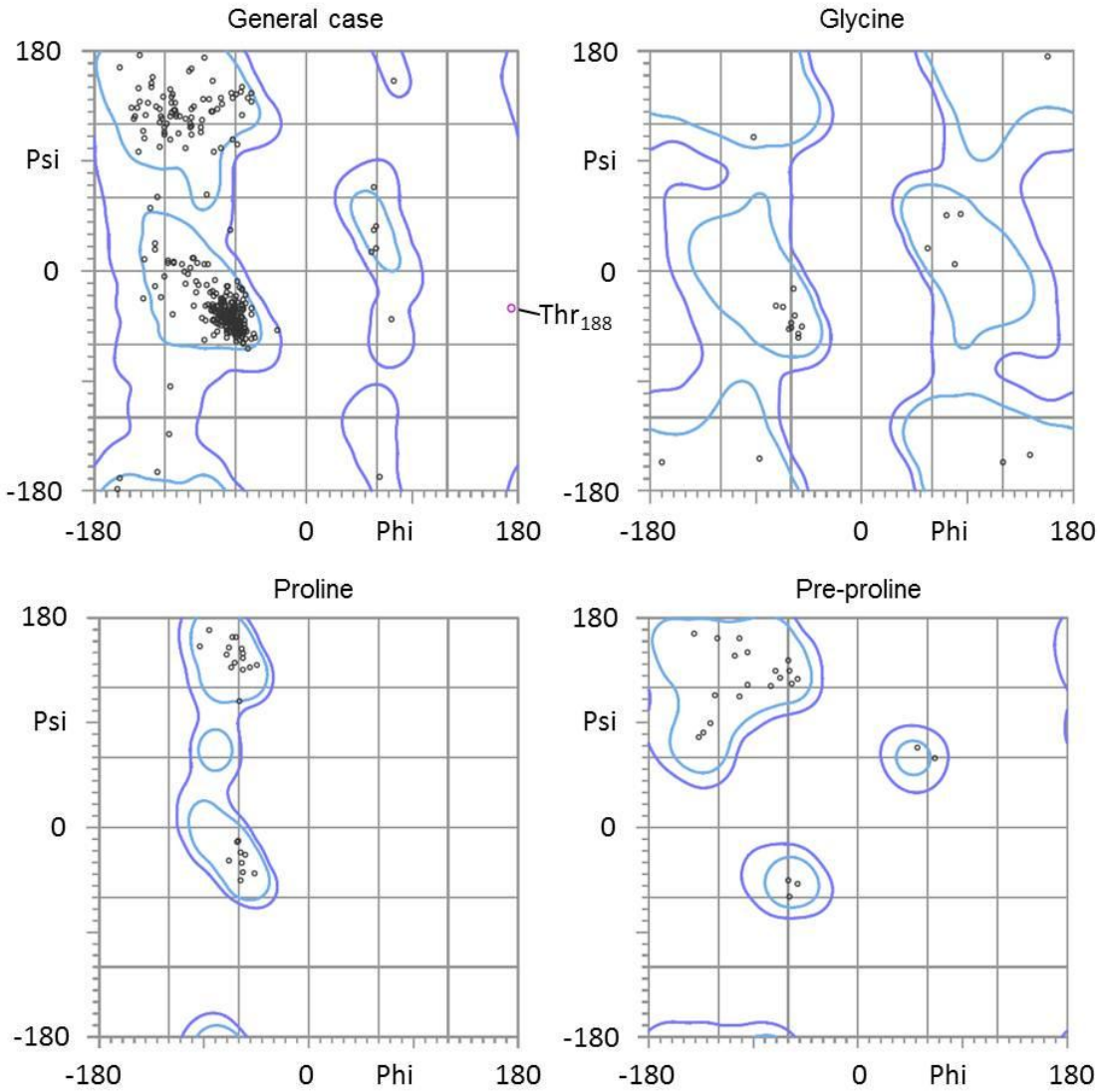


Figure 5-4: MolProbity Ramachandran analysis of Fragment 11 structure.

5.4.4 Interpretation of the Fragment 11 crystal structure

5.4.4.1 Overview

An overview of the Fragment 11 crystal structure is shown in Figure 5-5.

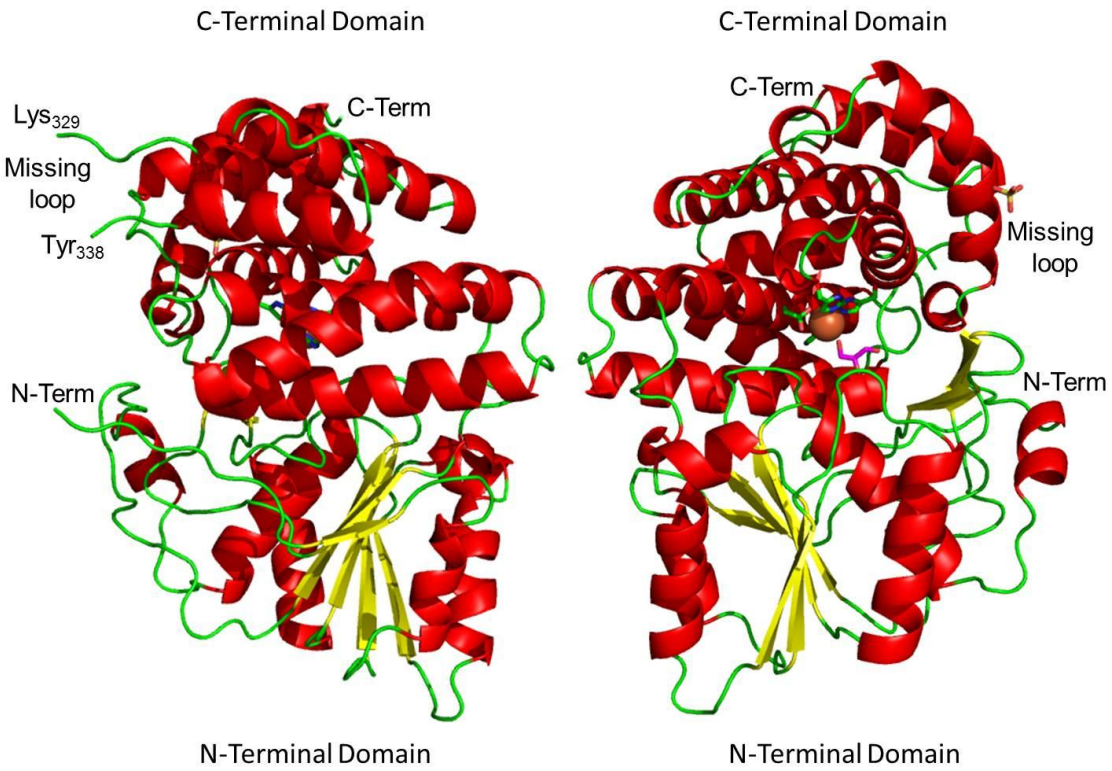


Figure 5-5: A cartoon view of the Fragment 11 crystal structure in front and back views. Red spirals = α-helices, yellow arrows = β-strands & Term = terminus. A metal ion (brown sphere), glycerol (purple) and a sulphate molecule (yellow) are also shown in the structure.

Fragment 11 is composed of two structural domains. The N-terminal domain has a 3-layer (αβα) sandwich or Rossmann fold architecture (CATH code: 3.40.50.1970) typical of NAD(P)$^+$ cofactor-binding domains (Rossmann et al. 1974). The C-terminal domain is α-helical with an Up-down Bundle architecture known as a dehydroquinate synthase-like α-domain (CATH code: 1.20.1090.10).

Poor electron density was observed for one loop region of the protein (F_{330}-E_{337}) so this was excluded from the structure. As shown in Figure 5-6, a gap in the unit cell packing can be observed around the missing loop region. This loop region was likely to be in multiple conformations and thus difficult to determine. No other region of the protein is occupying this space. This implies that the loop had not been cleaved but was hypermobile and not occupying one fixed conformation. Density for the N-terminal His-tag was also not observed in the data (MGSSHHHHHHSSGLVPRGSHMM). The last methionine in the missing tag is the first coding residue of the ADH domain of ADHE. The unstructured flexible nature of the additional tag means it is likely to be in multiple conformations and so could not be resolved from the data.

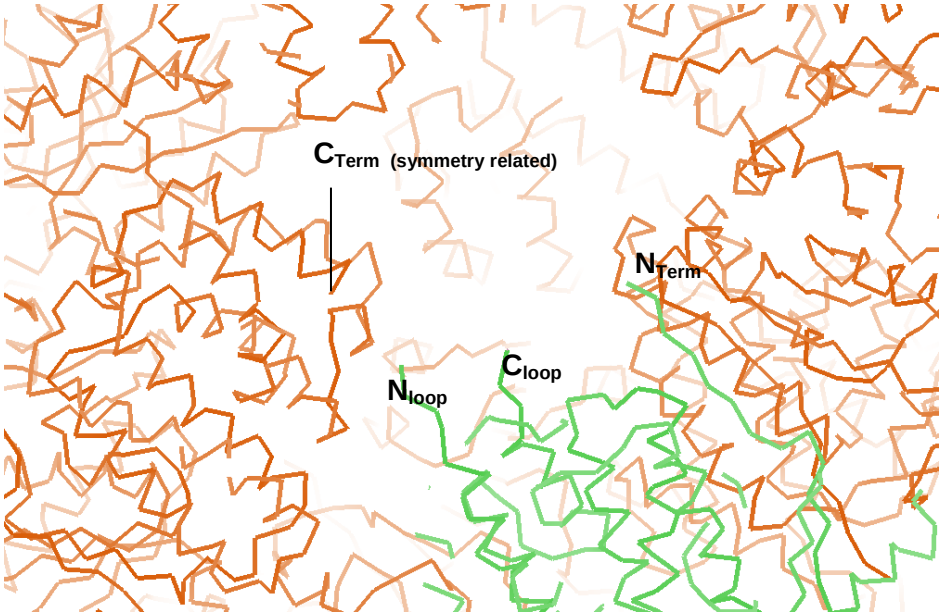


Figure 5-6: Cα traces of symmetry-related molecules (orange) surrounding the missing loop region in the Fragment 11 structure (green). Proximal symmetry-related termini are indicated as are the termini of the truncated loop.

Positive electron density was observed at the terminal S_γ of Cys_{409}; when the residue was changed to an S-hydroxycysteine during refinement, the density fit was much improved. The residue may be readily oxidised through radiation damage caused by exposure to high-energy synchrotron radiation. This modified residue has been observed in other high-resolution protein crystal structures within the PDB (292 entries (RCSB 2012)). The residue appears close to the glycerol in the active site but does not appear to form interactions with that molecule.

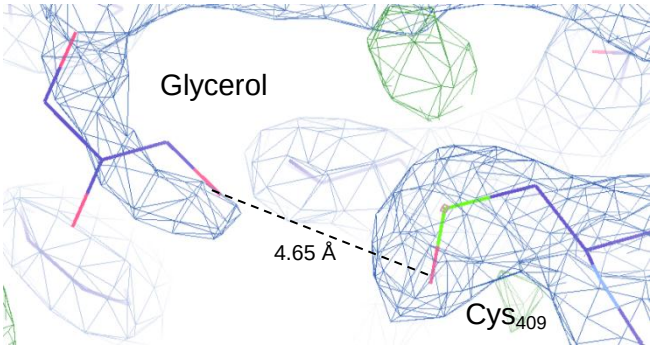


Figure 5-7: Electron density map for modified Cys_{409} (S-hydroxycysteine) showing the improved fit, lack of difference density and distance to the "glycerol" in the active site. Blue map = $2F_o$ - F_c (where $F_o = F_{observed}$, $F_c = F_{calculated}$, and F = structural factor) σ level = 1.02, negative (red) and positive (green) difference map ($F_o - F_c$) σ level = 3.0. Molecules are shown in stick form, with carbons in purple, and the glycerol is indicated.

5.4.4.2 Dimerization

During biochemical characterisation of the Fragment 11 protein (Chapter 4), it was shown to form dimers by gel filtration and DLS techniques. It would appear that a dimeric structure is also present within the unit cell of the crystals.

The EBI Protein Interfaces, Surfaces and Assemblies service (PISA) was used to evaluate the observed interactions between the protein monomers; this analysis predicts the most likely physiologically relevant interaction (Krissinel and Henrick 2007). Within the unit cell there were 3 possible interactions between Fragment 11 monomers, and these are shown in Figure 5-8. Each of these interactions was evaluated using PISA, which determined the likelihood of the particular interaction being involved in complex formation.

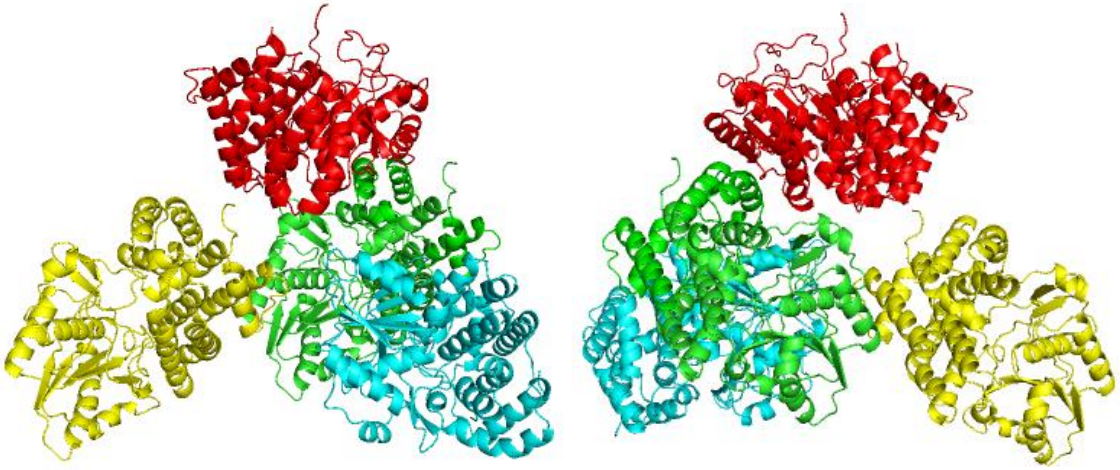


Figure 5-8: Cartoon diagram for the three interactions between Fragment 11 molecules within the crystal (front and back views). Green = molecule A, blue = interaction partner 1, red = interaction partner 2, and yellow = interaction partner 3.

Interaction #	# interfacing residues		Interface area (Å^2)	Δ^iG (kcal/mol)	Δ^iG P-value	#HB	#SB	CSS
	A	B						
1	49	50	1774.5	-26.9	0.019	20	6	0.563
2	23	24	844.5	-8.5	0.210	7	0	0.000
3	13	10	444.0	-0.6	0.697	3	1	0.000

Table 5-5: PISA analysis results summary. Δ^iG = solvation free energy gain upon interface formation (-ve = hydrophobic interface). P-value = probability that the decrease in Δ^iG through the interface is not random, where P>0.5 = non-specific and P<0.5 shows interface that may be considered interaction specific. HB = hydrogen bonds formed, SB = salt bridges formed. CSS = complexation significance score, which is a measure of interface significance in terms of complex formation (the higher the value between 0 and 1, the more significant the interface).

PISA analysis showed the most significant interface between the monomers was the interaction on the N-terminal side of the two proteins (interaction 1). The dimerization interface is shown in Figure 5-9 and key interaction-forming residues are shown in Table 5-6. Interaction 2 is coordinated by an SO_4^{2-} molecule (discussed in 5.4.4.7). This is likely to be a crystallisation artefact rather than a physiologically relevant interaction. The PISA scores for interaction 3 are too low to be considered relevant.

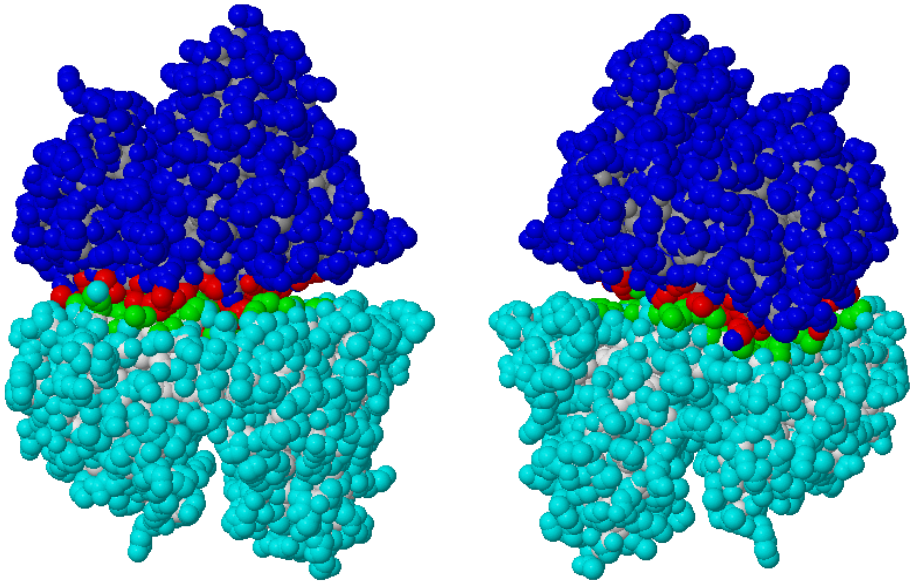


Figure 5-9: PISA interface space-filling diagram for interaction 1 (front and back views shown). Dark blue = molecule A, light blue = molecule B. Interface residues are shown in red for A and green for B.

Hydrogen Bond Forming Residues			
#	Molecule A	Distance (Å)	Molecule B
1	GLN 25 [N]	2.82	PHE 35 [O]
2	PHE 27 [N]	2.72	ILE 33 [O]
3	VAL 29 [N]	3.68	PRO 31 [O]
4	LYS 32 [NZ]	3.57	GLU 198 [OE1]
5	ILE 33 [N]	2.88	PHE 27 [O]
6	TYR 34 [OH]	2.84	ARG151 [O]
7	PHE 35 [N]	2.79	GLN 25 [O]
8	TYR 242 [OH]	2.85	ASN 276 [OD1]
9	LYS 273 [NZ]	2.96	ASP 241 [OD2]
10	ASN276 [ND2]	2.84	ASP 241 [OD1]
11	PHE 35 [O]	2.82	GLN 25 [N]
12	ILE 33 [O]	2.72	PHE 27 [N]
13	PRO 31 [O]	3.68	VAL 29 [N]
14	GLU 198 [OE1]	3.57	LYS 32 [NZ]
15	PHE 27 [O]	2.88	ILE 33 [N]
16	ARG 151 [O]	2.84	TYR 34 [OH]
17	GLN 25 [O]	2.79	PHE 35 [N]
18	ASN 276 [OD1]	2.85	TYR 242 [OH]
19	ASP 241 [OD2]	2.96	LYS 273 [NZ]
20	ASP 241 [OD1]	2.84	ASN 276 [ND2]
Salt Bridge Forming Residues			
1	LYS 32 [NZ]	3.57	GLU 198 [OE1]
2	LYS 273 [NZ]	3.93	ASP 241 [OD1]
3	LYS 273 [NZ]	2.96	ASP 241 [OD2]
4	GLU 198 [OE1]	3.57	LYS 32 [NZ]
5	ASP 241 [OD1]	3.93	LYS 273 [NZ]
6	ASP 241 [OD2]	2.96	LYS 273 [NZ]

Table 5-6: Summary of the PISA interface analysis result ([X] = atom identity from annotated residue, distance = distance between interacting atoms).

Schematic diagrams of the individual residues shown to be interacting to form the interface between the dimers are shown in Figure 5-10 and Figure 5-11.

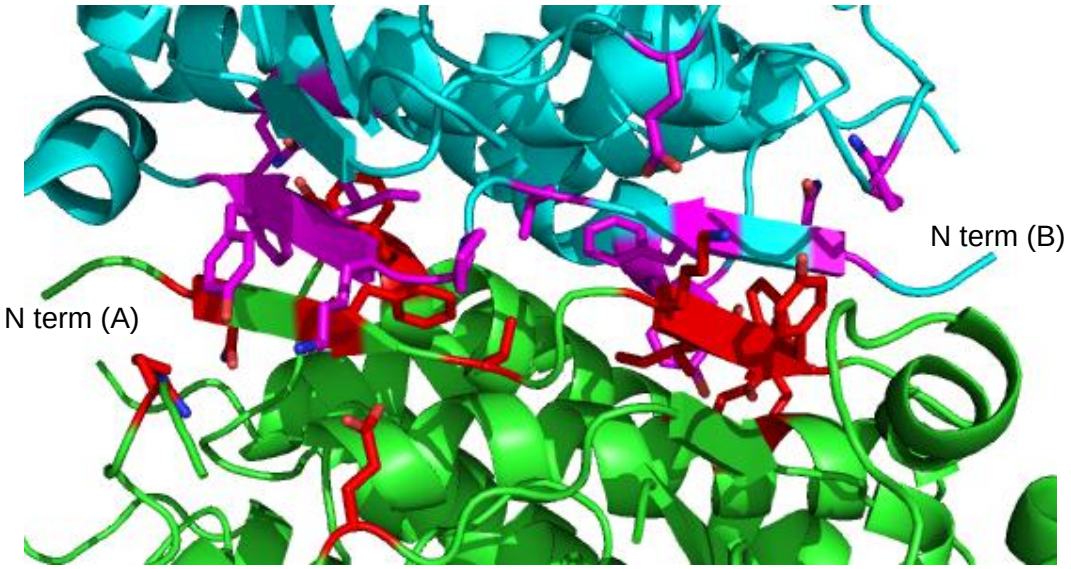


Figure 5-10: Cartoon diagram for the dimerization interactions between Fragment 11 molecules. Green = molecule A, blue = interaction partner 1 (B). Residues predicted by PISA to be involved in hydrogen bonding between the two molecules are coloured (red = A, purple = B). Some regions of the protein are excluded for clarity.

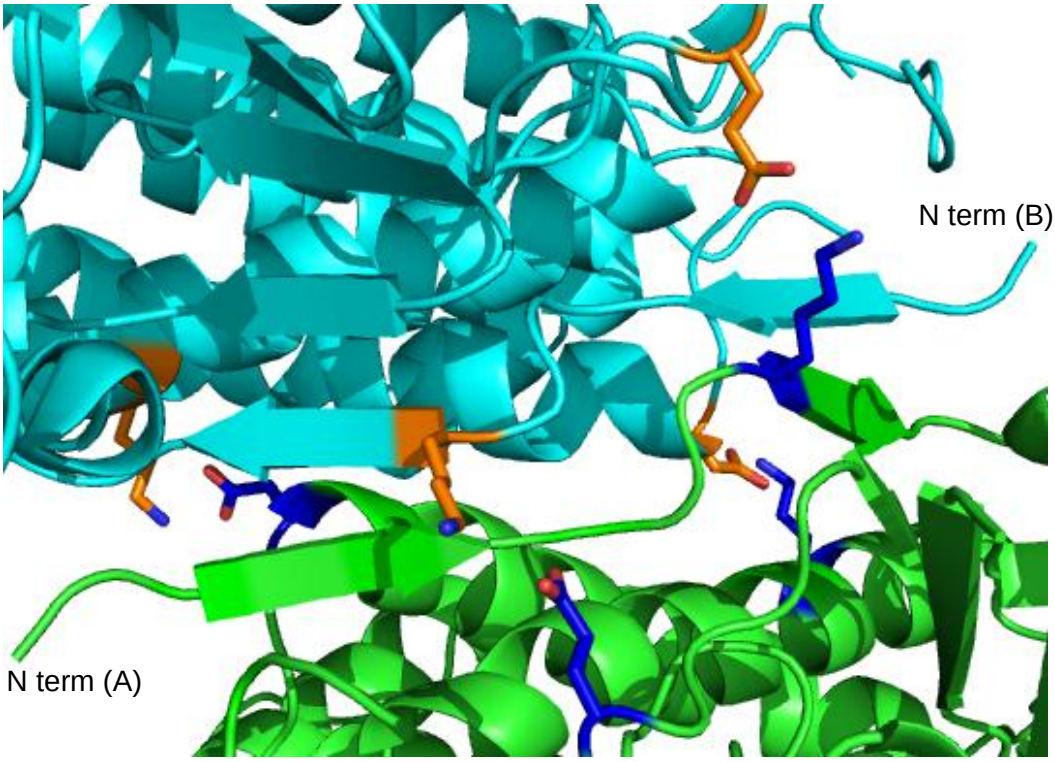


Figure 5-11: Cartoon diagram showing the salt bridges in the proposed dimerization interface between Fragment 11 molecules. Green = molecule A, blue = interaction partner 1 (B). Residues forming salt bridges between the two molecules predicted by PISA are coloured (blue = A, orange = B). Some regions of the protein are excluded for clarity.

Analysis showed that the physiologically relevant interaction between the monomers is the interaction on the N-terminal side of the proteins. This interaction is mediated through 20 hydrogen bonds and 6 salt bridges. The interface area for this interaction is much larger than for the other interactions and, according to the analysis, much more significant. This dimerization appears to orientate the proteins so the predicted active site clefts are pointing outwards away from one another as shown in Figure 5-12.

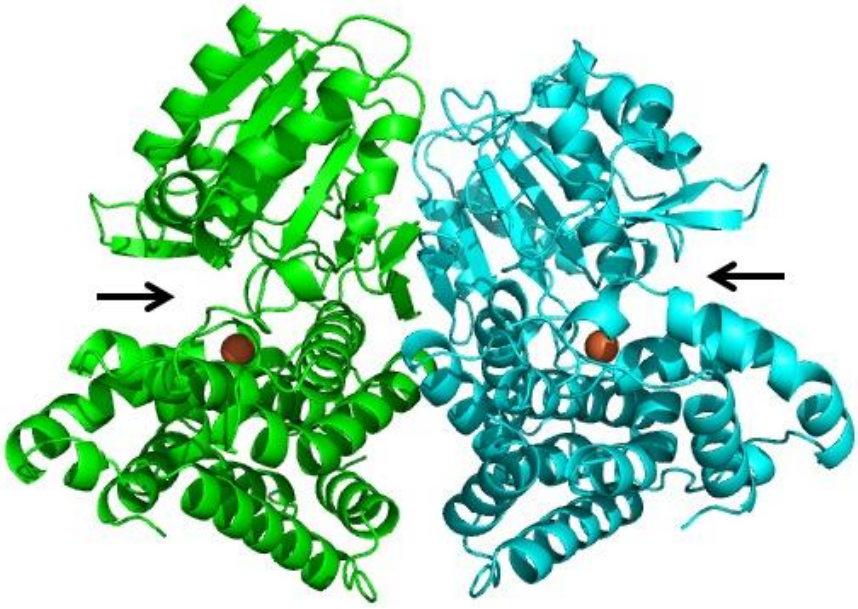


Figure 5-12: Cartoon diagram for the main interface between Fragment 11 molecules within the crystal. Green = molecule A, blue = molecule B and the active site metal ion is shown in brown. The predicted active site clefts are indicated with black arrows.

5.4.4.3 Visualisation by temperature factor (B-factor)

Visualising the structure by temperature factor (a measure of the mobility of the individual atoms present), allows mobile regions of the structure to be identified. In the following Figures an increasing width and "redness" (on a rainbow scale) of a region indicate increased mobility. It can be observed that the region around the predicted N-terminal dimerization interface has a limited mobility, whereas some of the outer loop regions are significantly more mobile. The side of the protein where the missing loop termini are situated appears to be more mobile than the opposite side.

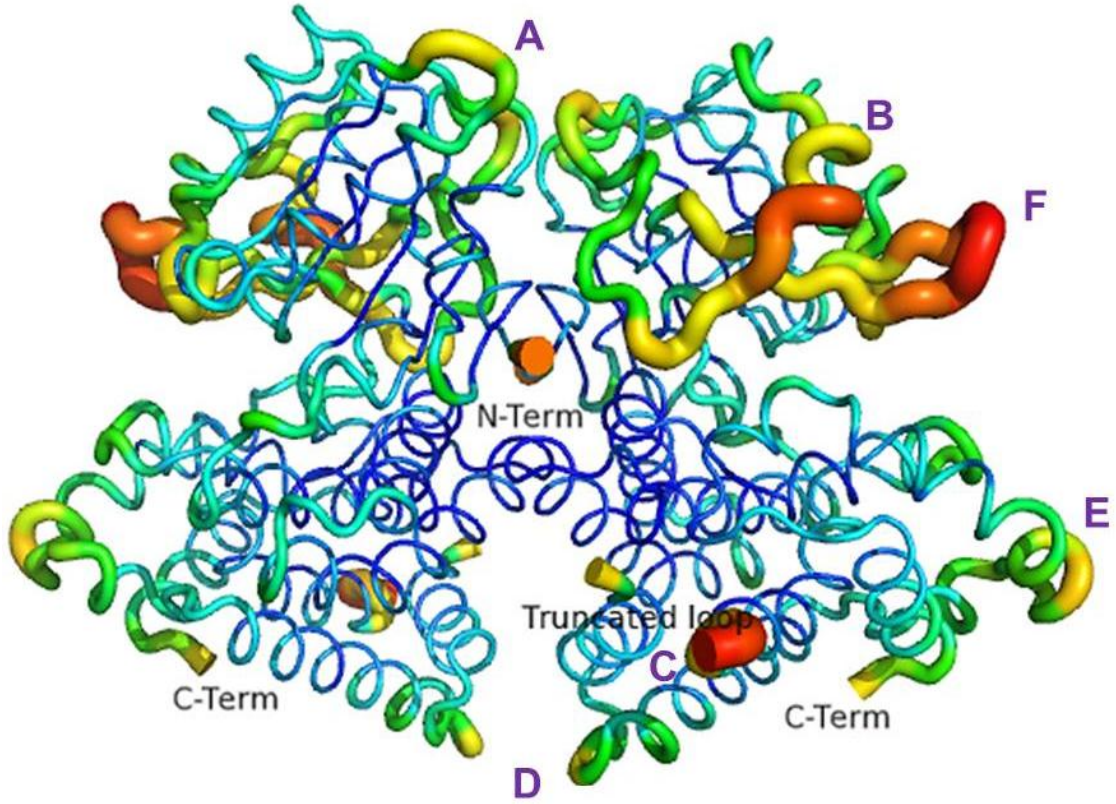


Figure 5-13: Fragment 11 visualised by temperature factor in a view showing the two structural domains making up each subunit. Wider and redder regions indicate increased mobility compared to thinner blue regions, which indicate limited mobility. The visible termini (Term) and truncated loops are indicated in the image. Purple A-F labels correspond to loops described in Figure 5-16.

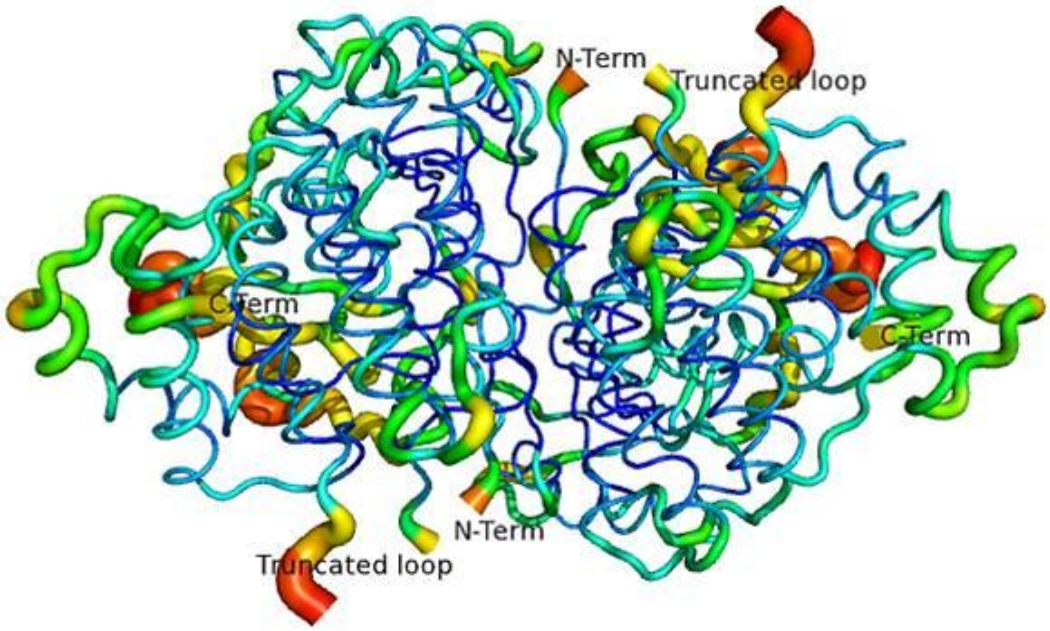


Figure 5-14: Fragment 11 visualised by temperature factor in a view looking down the 2-fold axis seen in Figure 5-13. Wider and redder regions indicate increased mobility compared to thinner blue regions, which indicate limited mobility. The termini (Term) and truncated loops are indicated in the image.

The most mobile region of the protein visible in the structure is loop F, (Ile_{183}-Pro_{193}), which has an average temperature factor of 93.6 $Å^2$ compared to the average of the whole molecule of 58.06 $Å^2$. Loop B (Thr_{135}-Lys_{161}) is the second most mobile region and has an average temperature factor of 81.0 $Å^2$. These loops are significantly more mobile than other regions of the protein. Fragment 11 corresponds only to the ADH domain of the ADHE protein. It is likely some stabilisation may be conferred by the N-terminal aldDH domain of the protein that is missing in Fragment 11. The relatively large difference observed between R and R_{free} for the structure is likely to be due to the high temperature factors observed indicating flexibility in these loop regions of the protein.

5.4.4.4 Structural alignment between Fragment 11 and homologous structures

Structural alignments between Fragment 11 and structures suggested by BALBES as good molecular replacement models were carried out; this process used the DaliLite pairwise protein comparison tool (EMBL) (Holm and Park 2000).

PDB code (chain ID)	Source organism	Protein description	Aligned residues	Seq ID	Z-score	RMSD (Å)
1RRM (B)	*E. coli*	Lactaldehyde reductase	366	35 %	47.5	2.3
3BFJ (A)	*Klebsiella pneumoniae*	1,3-propanediol oxidoreductase	377	38 %	54.3	1.8
1O2D (A)	*Thermotoga maritima*	Alcohol dehydrogenase, iron-containing	344	30 %	40.3	2.5

Table 5-7: Results of the structural alignment analysis using DaliLite. Z-Score describes the quality of the alignment (if Z-score >20, structures are considered homologous). RMSD is used as a measure of deviation in terms of distance of the Cαs in the aligned structures.

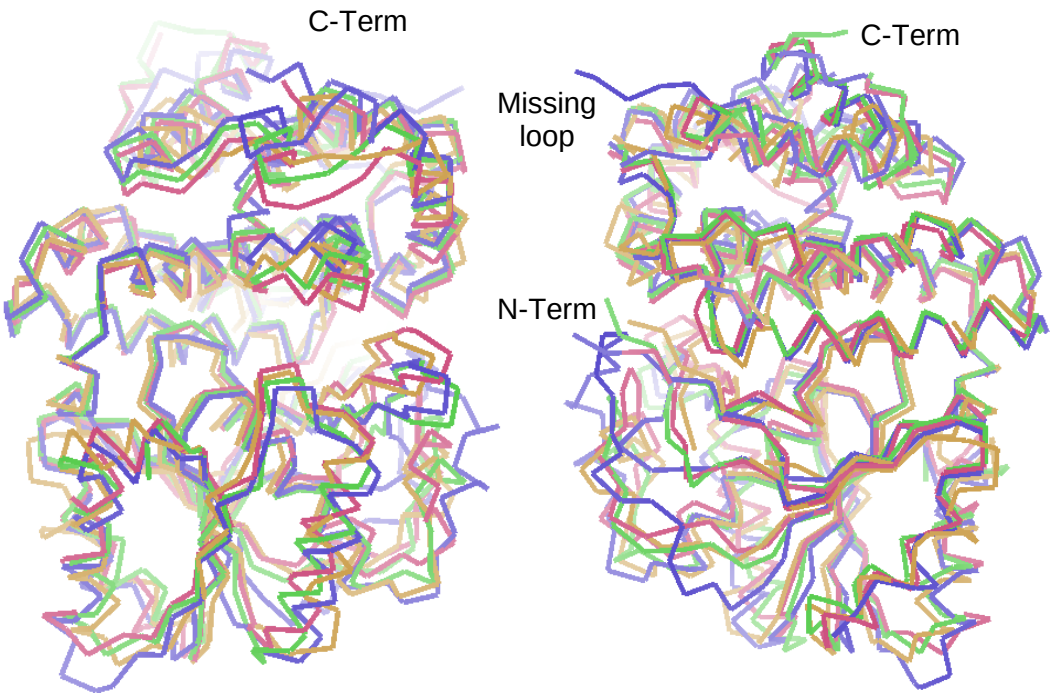


Figure 5-15: Cα traces of aligned Fragment 11 (purple), 1RRM (red), 3BFJ (green) & 1O2D (orange) "front and back" views shown. Visible termini are indicated as is the truncated loop of Fragment 11.

The interior secondary structure regions of the Fragment 11 protein align well with the other proteins in the alignments. The outer loops of the protein are the most divergent areas, and these are shown in detail in Figure 5-16. These loops also correspond to the most mobile regions of the protein as shown in Figure 5-13. The largest divergence between Fragment 11 and the other proteins would appear to be around the missing loop. It can be observed from the alignment that the other proteins have a much shorter loop in this position. Although loop F in Figure 5-16 is the most mobile in the structure, it appears to be relatively well conserved in the alignment between the homologous structures.

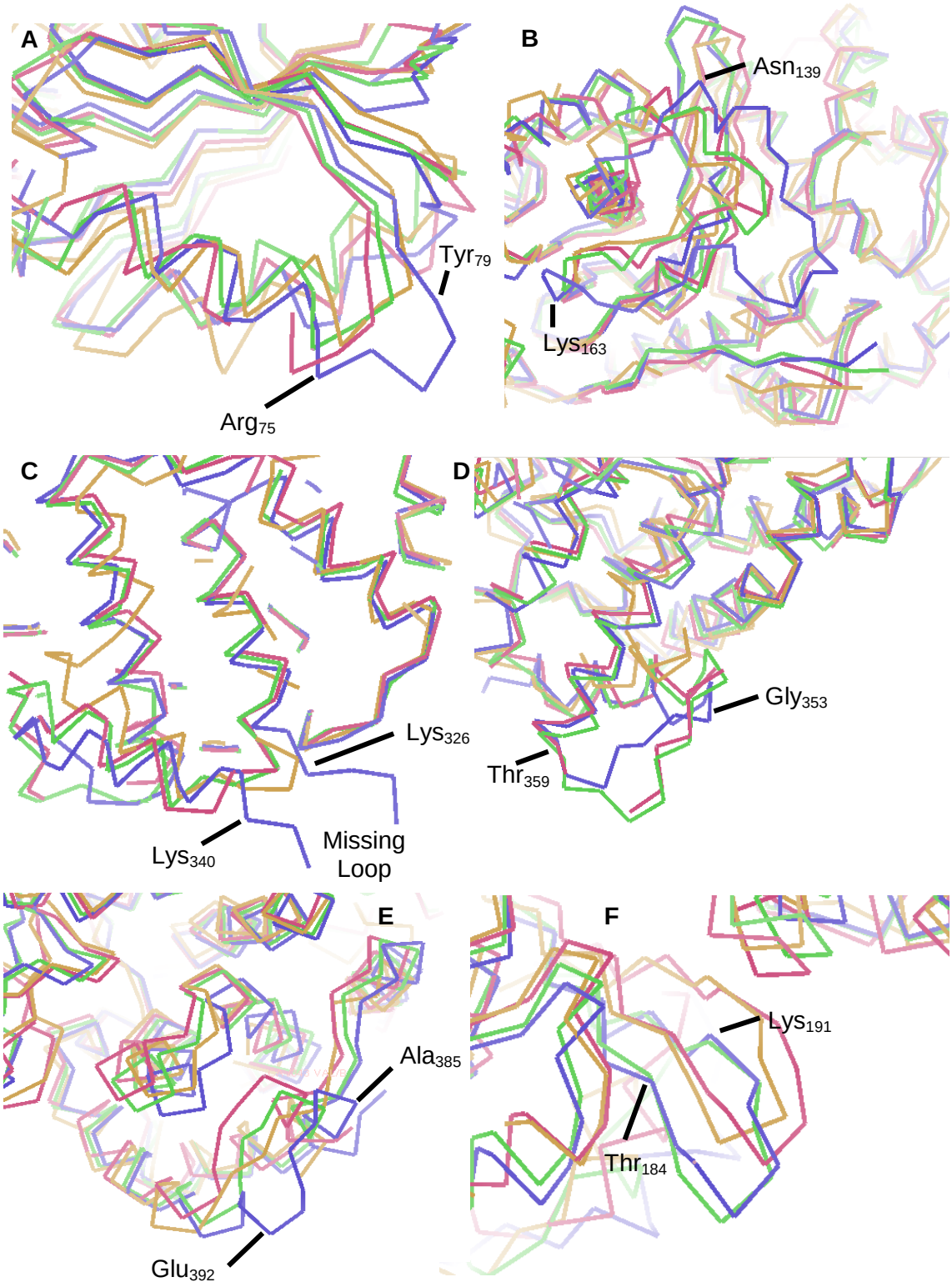


Figure 5-16: Cα traces of aligned Fragment 11 (purple), 1RRM (red), 3BFJ (green) & 1O2D (orange) with the major divergent loop regions annotated. Residues indicated are from the His-tagged Fragment 11 amino acid sequence. A-F labels correspond to loops annotated in Figure 5-13; F = the most mobile loop observed in structure.

Shorter loops can be observed in some regions for the aligned proteins compared to Fragment 11. As 1RRM (*E. coli*) and 3BFJ (*K. pneumonia*) both come from non-

thermophilic organisms, differences in these loops cannot be explained by differences in terms of growth temperature. The key difference between these proteins is the expression as individual domains as opposed to the fusion protein in the case of ADHE. This supports the hypothesised role of the aldDH domain in loop stabilisation of the ADH domain.

5.4.4.5 Metal ion coordination within the structure

Strong positive electron density was observed between an aspartic acid (Asp_{225}), three histidine residues (His_{229}, His_{294} and His_{308}) (which are all part of the α-helical domain of the protein) and a glycerol molecule. This density appeared to correspond to an octahedrally coordinated metal ion as shown in Figure 5-17.

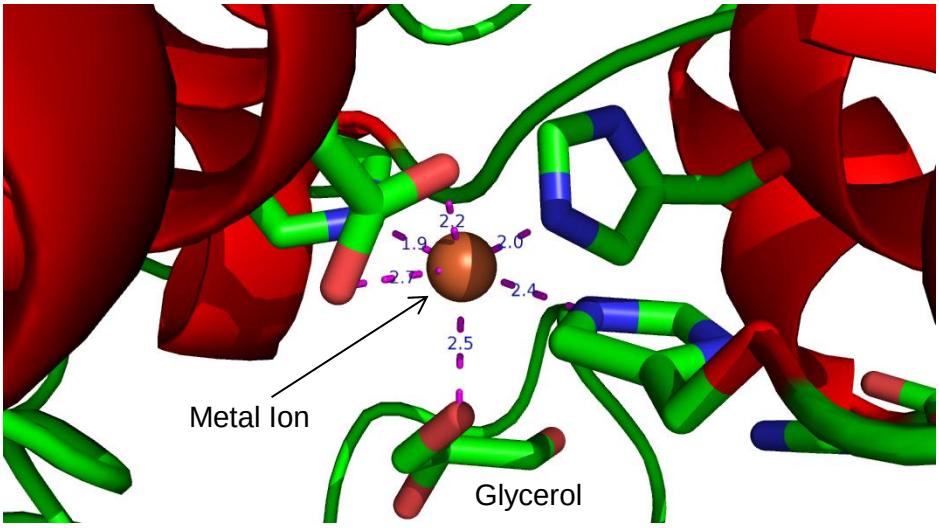


Figure 5-17: Cartoon of metal ion coordination in the Fragment 11 crystal structure. Red spirals = α-helices. The metal ion and glycerol molecule are indicated. Coordination distances (Å) are given in blue and their direction indicated by the purple dashed lines. Some regions of the structure are excluded for clarity.

Zn^{2+} and Fe^{2+} ions were fitted alternately into the observed density and the structure refined using Refmac5. Divalent metal ions were selected as the catalytic activity of the ADH domain of ADHE was shown to be stimulated by a variety of divalent metal ions. Zn^{2+} was used in the enzyme storage buffer used during crystallisation, but other proteins with this type of fold have been shown to contain an Fe^{2+} ion, so this metal ion was also considered.

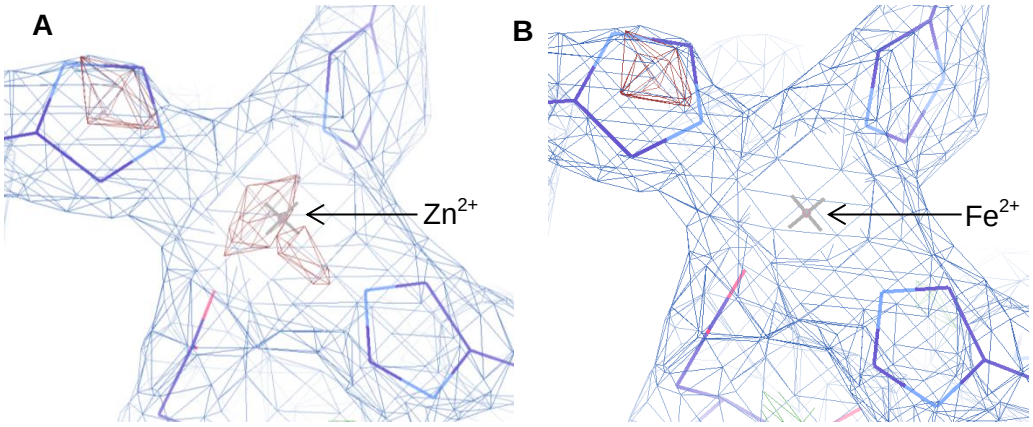


Figure 5-18: Electron density map for fitted metal ions. Blue map ($2F_o$ - F_c) σ level = 1.02, negative difference (red) and positive difference maps (green) (F_o – F_c), σ level = ±3.0. Amino acids are drawn in stick form with carbon atoms shown in purple. Grey cross = metal ion (Zn^{2+} in A, Fe^{2+} in B).

Negative electron density visible below σ level = -3.5 was observed when a Zn^{2+} ion was introduced. The Fe^{2+} ion fitted the observed density better with negative electron density only being visible below σ level = -3.0 (Figure 5-18). The additional electrons present in Zn^{2+} appeared to over-account for the density observed in the structure.

Metal ion coordination analysis of structures in the PDB showed that in catalytic sites both Zn^{2+} and Fe^{2+} are commonly coordinated by histidine residues (Dokmanic et al. 2008). Zn^{2+} was shown to be more common with coordination number 4 where the metal ion is in a structural site, whereas in a catalytic site both coordination number 5 and 6 are commonly observed. It should be noted that the number of structural zinc sites within the PDB appears higher than catalytic sites. Fe^{2+} tends to be coordinated with a number of 6 (Dokmanic et al. 2008). Although this does not allow the metal ion in Fragment 11 to be unambiguously identified, it does indicate that Fe^{2+} or Zn^{2+} are the most likely metal ions to be present in the active site of this protein.

5.4.4.6 Ion analysis

In an attempt to identify the metal ion present in the protein, ion analysis was carried out as described in Section 5.2.6 using Atomic Absorption Spectroscopy. Initially, this was carried out on a purified His-tagged Fragment 11 sample and a ratio of 2.5 zinc atoms : 1 polypeptide was observed. This was higher than would be expected as only one metal ion appeared in the Fragment 11 crystal structure.

It has been shown that His-tags on recombinant proteins can be involved in the binding of zinc ions (Evers et al. 2008) and therefore thrombin protease was used to

specifically remove the tag from Fragment 11 prior to re-analysis. Thrombin cleavage and re-purification was carried out as described in Section 5.2.7 (Figure 5-19). The samples were assayed to check activity was still present and, following dialysis into 50mM EPPS buffer pH 8.0, were re-analysed.

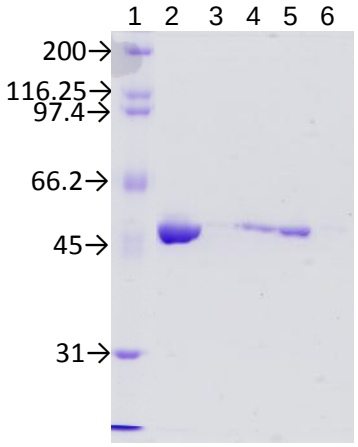


Figure 5-19: SDS-PAGE analysis of thrombin-cleaved Fragment 11 after metal affinity chromatography. 1 = markers (M_r/1000), 2 = load sample, 3 = flow through, 4-6 = wash steps where 4 = 0%, 5 = 5%, and 6 = 100% HIS-ELUTE buffer. The 0% sample was selected for dialysis. The M_r of cleaved Fragment 11 = 46,727.

Atomic Absorption Spectroscopy analysis revealed a ratio of 0.43 zinc atoms per protein monomer.

This experiment was repeated in an attempt to determine if iron was present within the structure. No iron was detected within the protein samples tested (thrombin-cleaved and un-cleaved), although the sensitivity of the instrument to iron solutions was significantly lower than for zinc and levels may have been below those required for detection.

Analysis by SEM-EDS for the cleaved and un-cleaved Fragment 11 samples (in buffer and H_2O) also did not detect any iron within the various samples tested. Low levels of zinc and nickel were detected in some of the samples. This may indicate that the protein is capable of scavenging a range of divalent metal ions.

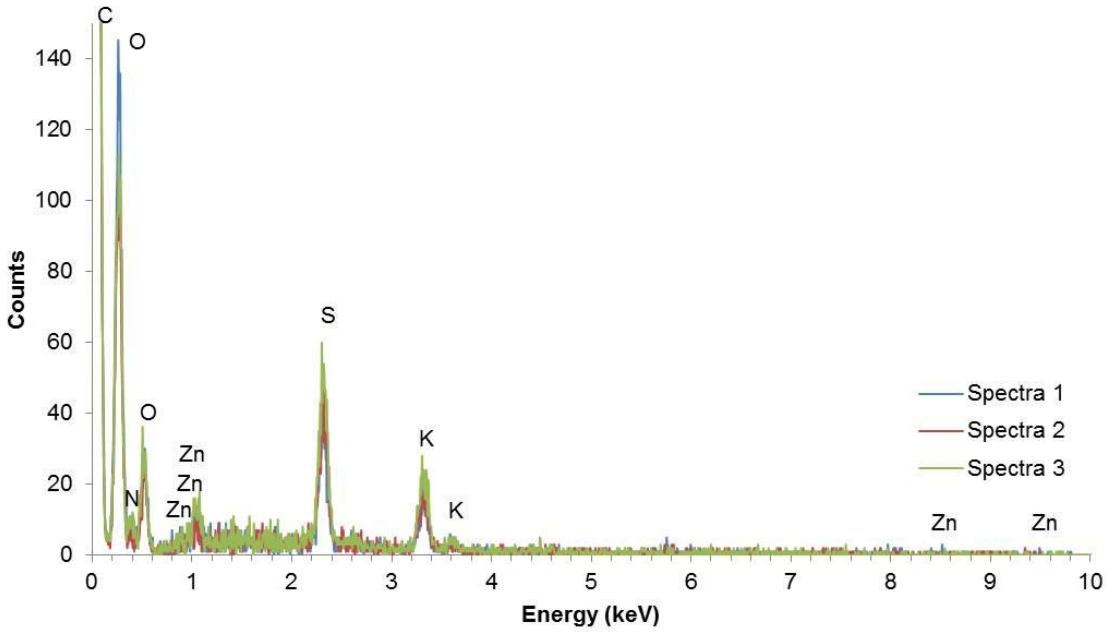


Figure 5-20: Example spectrum from SEM-EDS analysis of cleaved Fragment 11 in H_2O. Element peaks detected are labelled. Counts for carbon were up to 1300, but the graph is scaled to visualise less abundant peaks.

Fe^{2+} was not detected during analysis of the protein samples, whereas some Zn^{2+} was detected. Zn^{2+} was therefore modelled into the final structure of the protein. Ion analysis indicated less than 1 Zn^{2+} per polypeptide was present and during refinement a fully-occupied zinc atom appeared to over-account for the observed density. The occupancy of the zinc was therefore varied until a reasonable temperature factor for the metal ion was obtained. The average side chain B-factor is 61 $Å^2$ whereas the average for the metal coordinating atoms is 58 $Å^2$. A reasonable B-factor for the metal ion was determined to be 59 $Å^2$, which was observed when the occupancy of the zinc was reduced from 1 to 0.8. Partial occupancy by the Zn^{2+} ion fitted the observed density much better, with negative electron density only being visible below σ level = -2.7.

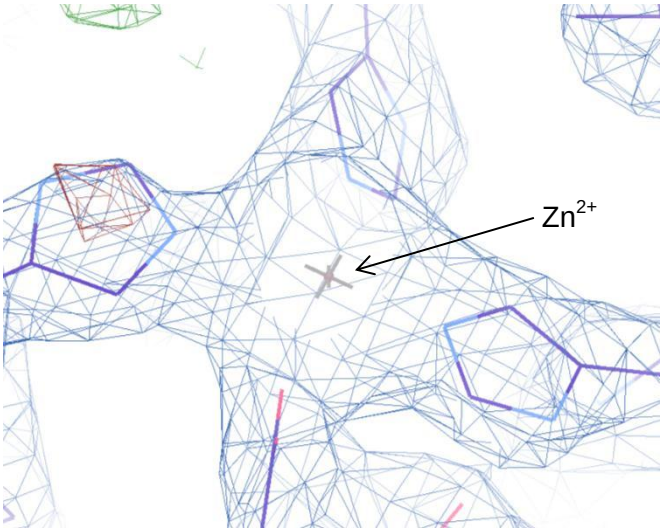


Figure 5-21: Electron density map for fitted Zn^{2+} ion. Blue map ($2F_o$ - F_c) σ level = 1.09, negative difference (red) and positive difference maps (green) (F_o – F_c), σ level = ±3.0. Amino acids are drawn in stick form with carbon atoms shown in purple. Grey cross = Zn^{2+} ion (occupancy = 0.8).

5.4.4.7 Other molecules observed in structure

Some regions of anomalous positive density were observed in the structure, with the highest located close to the coordinated metal ion. Initially this was modelled by two water molecules and then a fragment of PEG; however, glycerol (which was used as a cryo-protectant) appeared to fit best when refined into this density. The molecule appears to be interacting with the active-site metal ion (Figure 5-22).

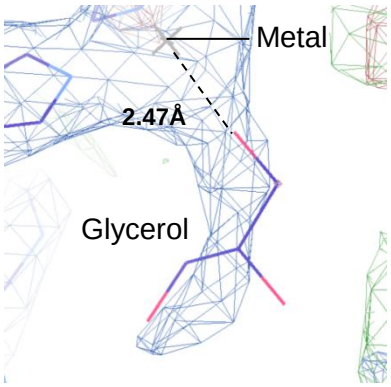


Figure 5-22: Electron density map for fitted glycerol. Blue map ($2F_o$ - F_c) σ level = 1.02, negative difference (red) and positive difference maps (green) (F_o – F_c), σ level = ±3.0. Amino acids and glycerol are drawn in stick form, with carbon atoms shown in purple. Grey cross = Zn^{2+} ion. Environment distances for the glycerol are shown.

Although the density fit of the glycerol is not ideal, it appeared to be the best fit of the various molecules tested. An interaction between a terminal OH group and the active-site metal ion was observed for both the PEG and glycerol molecules when refined into the structure. The terminal alcohol group in glycerol is similar to that of ethanol (Figure 5-23) and so was considered to be a product mimic for this ADH domain. This gives an

indication as to the location of the active site of the protein between the Rossmann fold and the α-helical domain of the ADH.

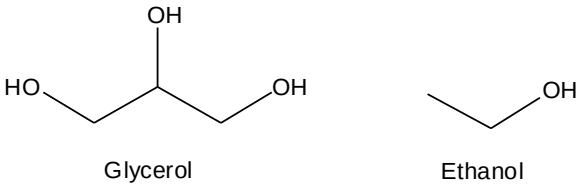


Figure 5-23: Structures of glycerol and ethanol.

A sulphate ion (present in the crystallisation buffer) was refined into some positive difference density next to Lys_{396}, at the interface between the protein and one of the symmetry-related molecules (interaction partner 2 in Figure 5-8). This density was too large to correspond to a water molecule and did not appear to be part of the protein itself. The sulphate ion appears to be interacting with the positively-charged side chains of two lysines (Lys_{396} & Lys_{399}) and two arginines (Arg_{75} & Arg_{74}) from a symmetry-related molecule.

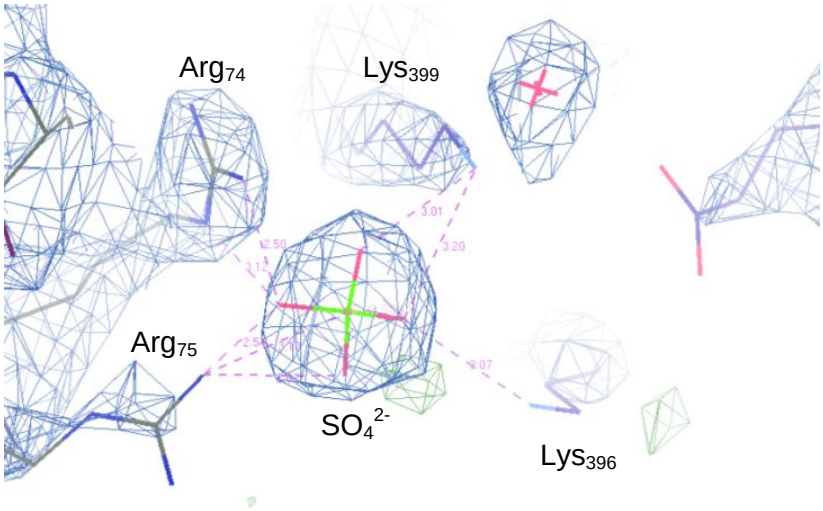


Figure 5-24: Electron density map for fitted SO_4^{2-}. Blue map ($2F_o$ - F_c) σ level = 1.02, negative difference (red) and positive difference maps (green) (F_o – F_c), σ level = ±3.0. Molecules are drawn as sticks with carbons in purple and in grey in the symmetry-related molecule. Environment distances for SO_4^{2-} are shown (Å).

5.4.4.8 NAD^+ binding

Two predicted $NAD(P)^+$ binding motifs (GXGS) are annotated within the Fragment 11 amino acid sequence. These motifs appear in the loop regions of the Rossmann fold, facing the α-helical domain of the protein that includes the previously-annotated metal ion binding motif (Chapter 1).

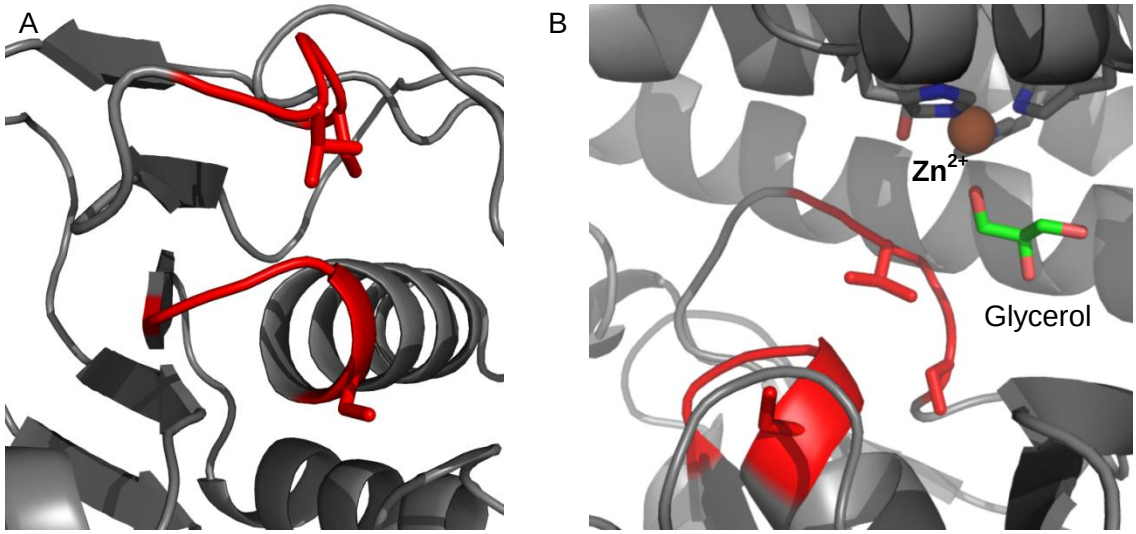


Figure 5-25: Cartoon of NAD(P)$^+$ binding motif regions in the Fragment 11 crystal structure, shown in top (A) and side (B) views. Spirals = α-helices, arrows = β-strands & red regions = GXGS motifs. The Zn^{2+} and glycerol present in the structure are indicated. Some regions are excluded for clarity.

Attempts to obtain a Fragment 11 structure with cofactor bound were made during data collection; the crystals were soaked in a cryo-protectant-well solution containing 0.2 or 0.04 mM NAD$^+$. Immediately upon introduction of the crystal to the cofactor-cryo-protectant mixture, the crystals were observed to crack and no longer diffracted X-rays sufficiently. This is likely to be due to a significant conformational change induced by the presence of the cofactor.

The 1O2D structure (Fe-ADH from *Thermotoga maritima*) contains an NADP$^+$ molecule close to the predicted active site of the protein. As the structure of Fragment 11 aligns reasonably well with 1O2D, it is possible to superimpose the NADP$^+$ from 1O2D onto Fragment 11 following structural alignment to provide an estimate of where the NAD(P)$^+$ cofactor may bind in this structure. These cofactors differ only with respect to the presence/absence of a 2' phosphate group of the ribose ring closest to the adenine ring of the molecule.

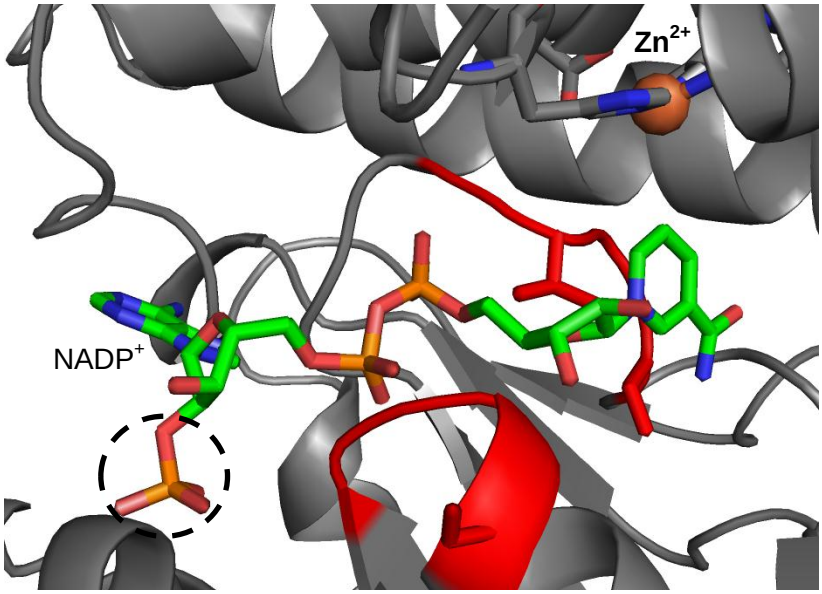


Figure 5-26: Cartoon of the $NAD(P)^+$ binding motif regions in the Fragment 11 crystal structure with a model $NADP^+$ superimposed. Spirals = α-helices, arrows = β-strands & red regions = GXGS motifs. The Zn^{2+} and superimposed $NADP^+$ are indicated. The 2' additional phosphate group present in $NADP^+$ is highlighted with a dashed circle. Some regions of the Fragment 11 structure are excluded for clarity.

Modelling the $NADP^+$ from 1O2D provides an indication of the residues that may be interacting with $NAD(P)^+$ in the active site of this domain. It should be noted that, without a structure in the presence of $NAD(P)^+$, these predicted interactions cannot be determined accurately. The CONTACT programme (CCP4i) was used to analyse the predicted interactions between Fragment 11 and the $NADP^+$. Hydrogen bonding and hydrophobic interactions were observed, and these interactions are compared to those observed in 1O2D in the following Tables and Figures. As the ADH domain of ADHE is more specific for NADH than NADPH (Chapter 3), interactions between the protein and the 2' additional phosphate group present in $NADP^+$ were not considered during the comparison.

NADP+ Atom	Frag 11 Residue	Frag 11 Atom	Distance Å	1O2D Residue	1O2D Atom	Distance Å
N1A	Thr 213	OG1	2.81	Ser 177	OG	2.69
	Val 214	N	3.23			
N6A	Thr 169	O	2.96	Thr 136	O	2.82
	Phe 210	O	2.52	Tyr 174	O	2.85
				Ser 177	OG	3.28
N7A	Thr 169	OG1	3.16	Thr 136	OG1	2.93
	Thr 169	O	3.07			
N7N	Asp 122	OD1	3.98	Asp 100	OD1	2.76
	Thr 178	OG1	3.37	Tyr 147	O	3.07
	Phe 180	O	3.93			
	Ser 175	OG	2.8			
O1A				Ser 97	OG	2.68
O1N	Thr 170	OG1	3.07	Gly 96	N	2.93
	Thr 173	OG1	3.24	Thr 137	OG1	2.65
O2A	Gly 117	N	3.22	Ser 97	N	3.03
	Gly 118	N	2.87			
O2D				Lys 157	NZ	2.81
O3D				Lys 157	NZ	3
				Asn 69	OD1	3.28
O4D	Thr 173	OG1	3.07			
O5D	Gly 118	N	3.02	Gly 96	N	3.12
	Thr 173	OG1	2.81			
O7N	Thr 178	OG1	3.03	Thr 145	OG1	3.01

Table 5-8: Comparison of $NADP^+$ hydrogen bonding residues and distances for Fragment 11 and 1O2D crystal structures. $NADP^+$ atoms missing interactions are highlighted in red. Fragment 11 data based on CONTACT analysis (CCP4i); 1O2D data taken from the associated publication (Schwarzenbacher et al. 2004).

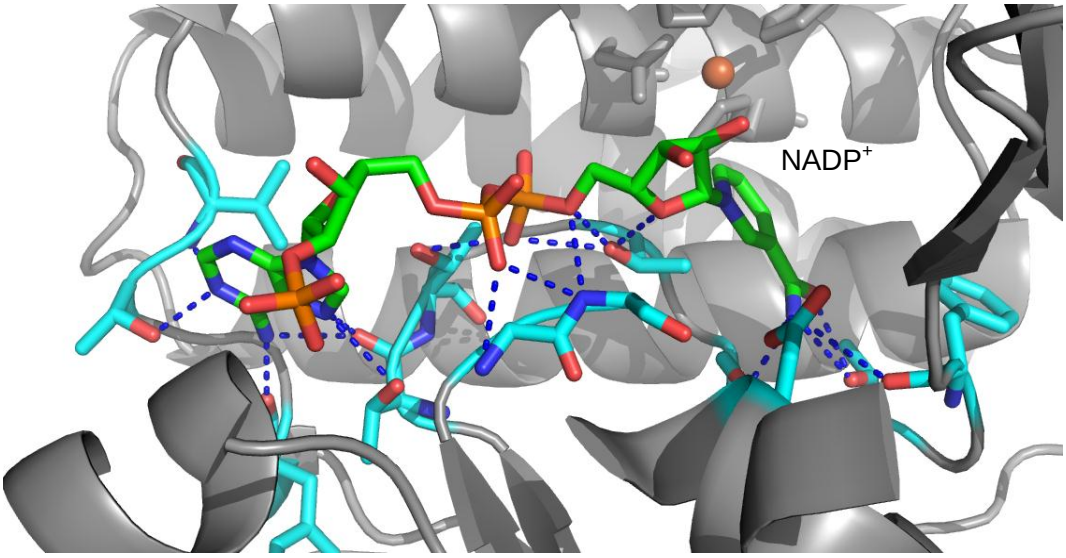


Figure 5-27: Cartoon of predicted hydrogen bonding between Fragment 11 and $NADP^+$ (from 1O2D) view shows the interface between the N and C terminal domains. Spirals = α-helices, arrows = β-strands, blue residues = hydrogen bonding residues, and blue dashes = hydrogen bonds. Some regions of the structure are excluded for clarity.

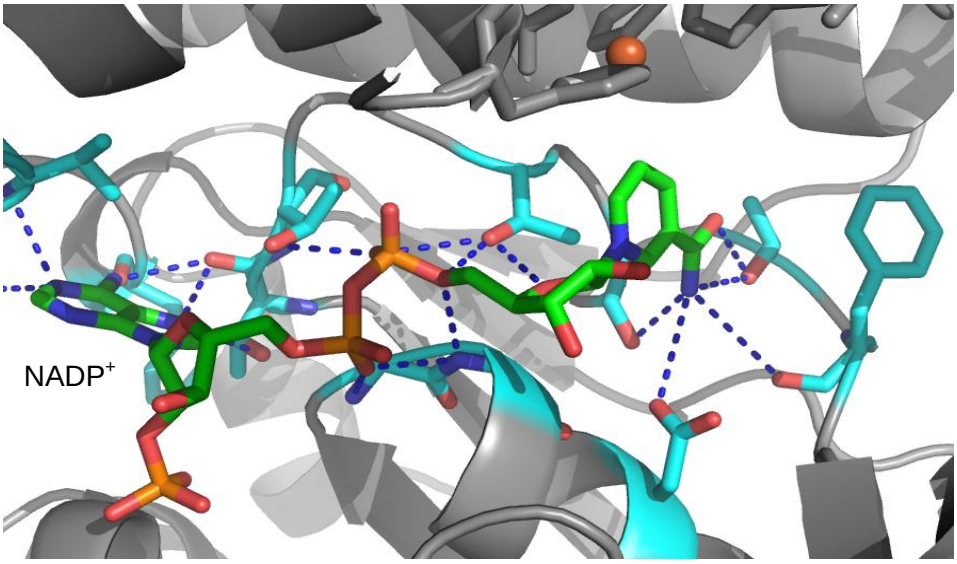


Figure 5-28: Cartoon of predicted hydrogen bonding between Fragment 11 and $NADP^+$ (from 1O2D) view shows $NADP^+$ from C-terminal domain side. Spirals = α-helices, arrows = β-strands, blue residues = hydrogen bonding residues, and blue dashes = hydrogen bonds. Some regions of the structure are excluded for clarity.

Most of the key hydrogen bonding interactions in the 1O2D structure (Schwarzenbacher et al. 2004) are conserved in Fragment 11, although, three interactions with the C-terminal domain are missing. It appears that the $NADP^+$ molecule is not close enough to the C-terminal domain of Fragment 11 to interact in the model. The Rossmann fold is the major $NAD(P)^+$ binding structural motif so these interactions would be expected to be highly conserved. Cofactor-induced structural changes are not predicted in the model, yet may have been the cause of crystal deterioration on soaking with NAD^+, so the C-terminal domain may be closer to the N-terminal domain when the cofactor is present in the protein.

$NADP^+$ Ring	**Frag 11 Residue**	**1O2D Residue**
Nicotinamide	Thr 173 His 229 Phe 285	Thr 140 Ile 149 His 193 His 270
Ribose 1		Thr 140
Ribose 2		Glu 68
Adenine	Met 60 Leu 63 Val 214	Met 178 Leu 182

Table 5-9: Comparison of residues interacting hydrophobically with $NADP^+$ in the Fragment 11 model and the 1O2D crystal structure. The missing interactions between the enzyme and the ribose rings of $NADP^+$ are highlighted in red. Fragment 11 data are based on CONTACT analysis (CCP4i), while 1O2D data are taken from Schwarzenbacher et al. (2004).

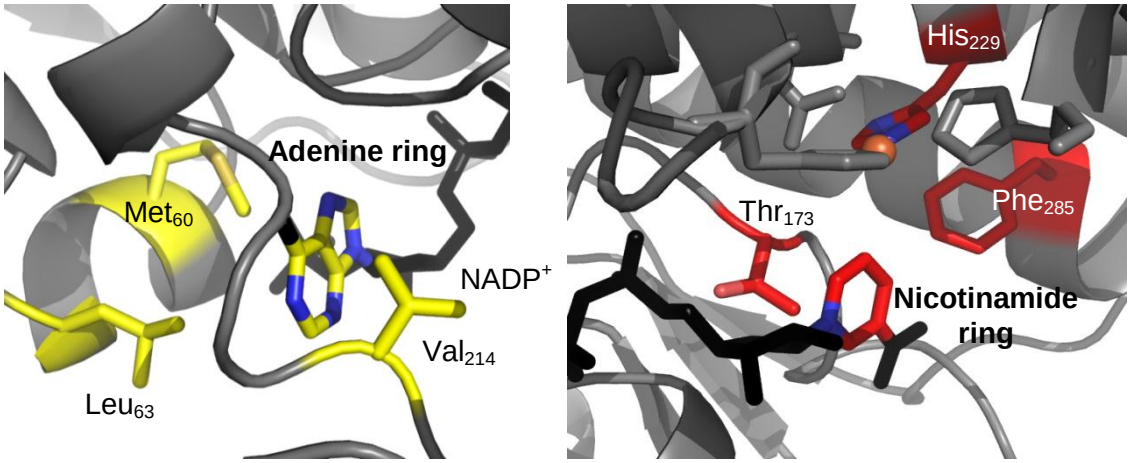


Figure 5-29: Cartoon of predicted hydrophobic interaction forming residues between Fragment 11 and $NADP^+$ (from 1O2D). Spirals = α-helices, arrows = β-strands, yellow regions = hydrophobically interacting residues with the adenine ring of $NADP^+$, and red regions = hydrophobically interacting residues with the nicotinamide ring of $NADP^+$.

Both the adenine and nicotinamide rings appear to be forming hydrophobic interactions with Fragment 11 residues in a similar position to those observed in 1O2D.

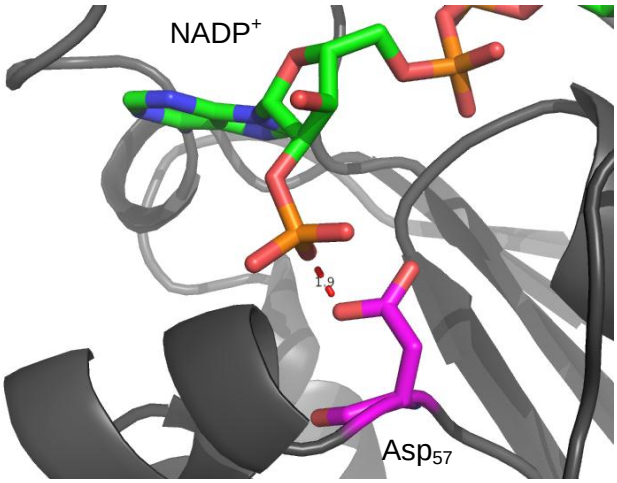


Figure 5-30: Cartoon of 2' phosphate of $NADP^+$ clashing with Asp_{57} in the Fragment 11 crystal structure. Spirals = α-helices, arrows = β-strands & purple residue = Asp_{57}. The superimposed $NADP^+$ is indicated. The distance between the Asp side-chain and the 2' phosphate group of $NADP^+$ is indicated by a red dashed line (1.9 Å). Some regions of the structure are excluded for clarity.

When the additional 2' phosphate (absent in NAD^+) was not considered, the $NADP^+$ cofactor fitted reasonably well into the structure of Fragment 11. However if $NADP^+$ did attempt to bind, Asp_{57} in Fragment 11 would clash with the 2' phosphate group of the ribose ring sterically and electrostatically (both negatively charged). In NAD^+ the diol group of the ribose ring may be stabilised through hydrogen bonding interactions with this residue. Key interactions were also missing in the Fragment 11 model with $NADP^+$

around the additional phosphate (coordinated by Ser_{38} & Ser_{39} in 1O2D). The ADH domain of ADHE catalyses the conversion of acetaldehyde to ethanol in the presence of NADPH at a much lower level (~70% lower activity) than with NADH; it is suggested that differences in these residues may be important in distinguishing the nicotinamide cofactors. Significant structural movements may need to occur in the region surrounding Asp_{57} to permit catalysis with NADPH.

5.4.5 N-terminal domain modelling work

Attempts to structurally resolve the ADHE protein and the aldDH domain were unsuccessful. *In silico* modelling work was carried out in an effort to predict the possible interactions between the aldDH domain and the ADH domain of the ADHE protein. Model creation, evaluation and docking was carried out by Dr Susan Crennell (University of Bath), but is reproduced here with permission.

5.4.5.1 Generation of a homology model of the aldDH domain

A homology model of the aldDH domain of the *G. thermoglucosidasius* ADHE protein was generated using the MODELLER programme (Sali and Blundell 1993). This was based on two homologous structures from within the PDB, 3MY7 (an aldDH domain of an ADHE from *Vibrio parahaemolyticus*) and 3K9D (a probable aldDH from *Listeria monocytogenes*). Sequence alignment scores for the proteins with the N-terminal domain of ADHE are shown in Table 5-10. The MODELLER programme generated a homology model of the aldDH domain of ADHE based on structure-based sequence alignments between the proteins; a structural alignment of this model and the two homologous structures is shown in Figure 5-31.

Protein sequence	% Identity	% Similarity	% Gaps
3MY7	46	67	4
3K9D	41	63	4

Table 5-10: Amino acid sequence alignment scores for residues 1-458 of ADHE with the two proteins used for modelling the aldDH domain.

The N-terminal region of the modelled aldDH domain of ADHE appeared to be unstructured compared to the homologous proteins used to generate the model. The model was truncated to Val_{10} for the subsequent docking work to prevent interference by this unstructured region. The C-terminal region of the protein appeared ordered so was not truncated.

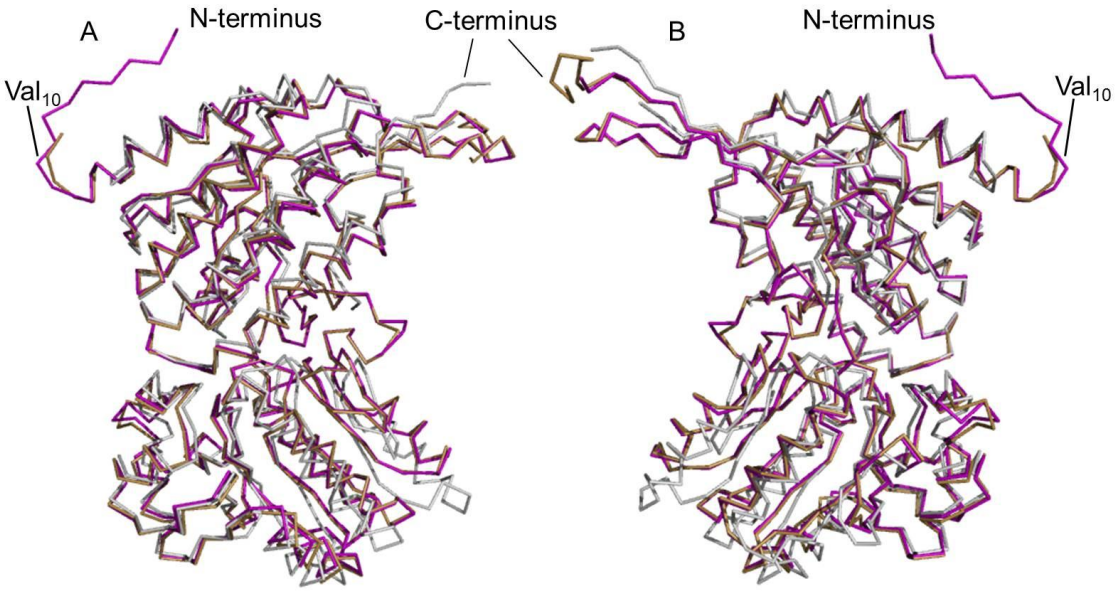


Figure 5-31: Cα traces of the aligned model aldDH domain (purple), 3MY7 (grey) and 3K9D (yellow). A = front view; B = back view. The termini of the proteins are indicated as is Val_{10} in the modelled aldDH domain.

The protein was modelled as a dimer as observed for the proteins of a similar fold, with the C-terminal regions interacting with the interaction partner as shown in Figure 5-32.

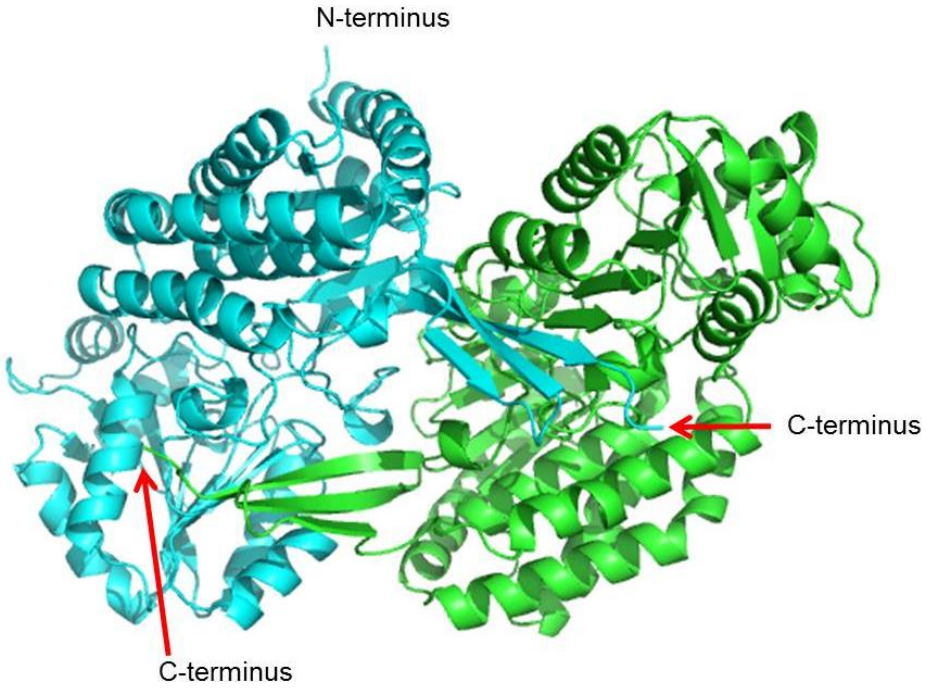


Figure 5-32: Cartoon diagram of the modelled dimeric aldDH. Green = molecule A, blue = molecule B. Visible termini in the Figure are indicated.

Two structural domains appear to be present in the aldDH domain of ADHE. The CATH assignment of these domains was assessed using the CATH structural

comparison tool. The N-terminal domain is part of the 3.40.605.10 superfamily (aldehyde dehydrogenase; chain A, domain 1) and the C-terminal domain is part of the 3.40.309.10 superfamily (aldehyde dehydrogenase; chain A, domain 2). The two domains both have 3-layer (αβα) sandwich topologies.

Evaluation of the model in comparison to the two homologous proteins used to create it (Figure 5-33) showed few significant differences in terms of the energy of the residues within the protein. Any differences that were observed were not above the maximum peaks observed for the 3K9D structure, suggesting the model was of reasonable quality.

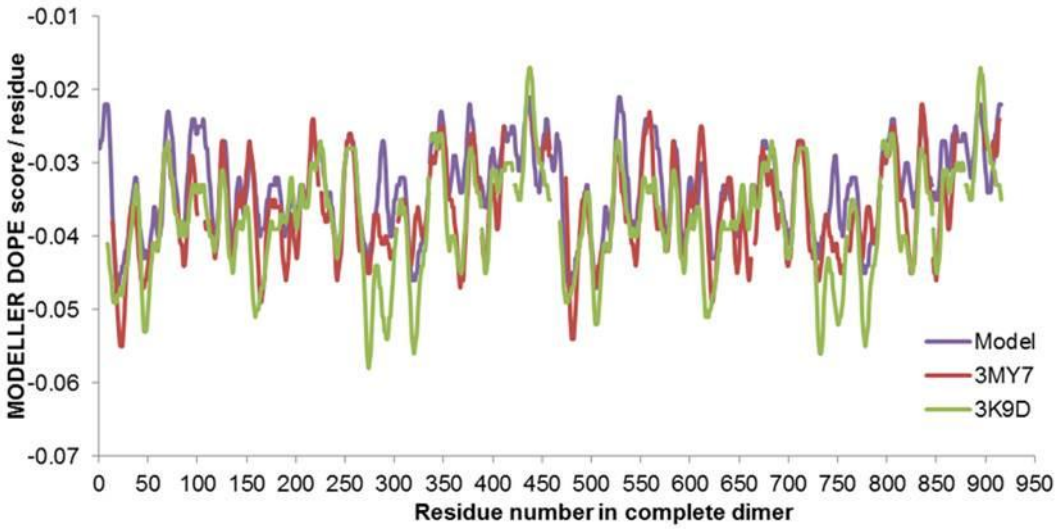


Figure 5-33: Modeller evaluation plot of Discrete Optimized Protein Energy (DOPE) score per residue for the aldDH model.

The main chain of the model was also evaluated using MolProbity; 891/894 residues were in Ramachandran allowed positions, leaving 3 residues as Ramachandran outliers. This shows that a majority of the main chain is modelled in an energetically favourable conformation. This evaluation suggested that the aldDH model structure was suitable for use in investigation of protein docking.

5.4.5.2 Docking of the aldDH model and Fragment 11

The aldDH model dimer was “docked” with the Fragment 11 dimer using three different protein interaction prediction programmes. No restrictions were imposed on the modelled interactions in terms of termini locations, or interaction face in an attempt to elucidate any biologically relevant interactions without bias. The three programmes used were ClusPro (Comeau et al. 2004), Hex (Ritchie and Kemp 1999) and ZDOCK

(Sali and Blundell 1993). The top-ranked interactions for each of the programmes are shown in Figure 5-34 and Figure 5-35.

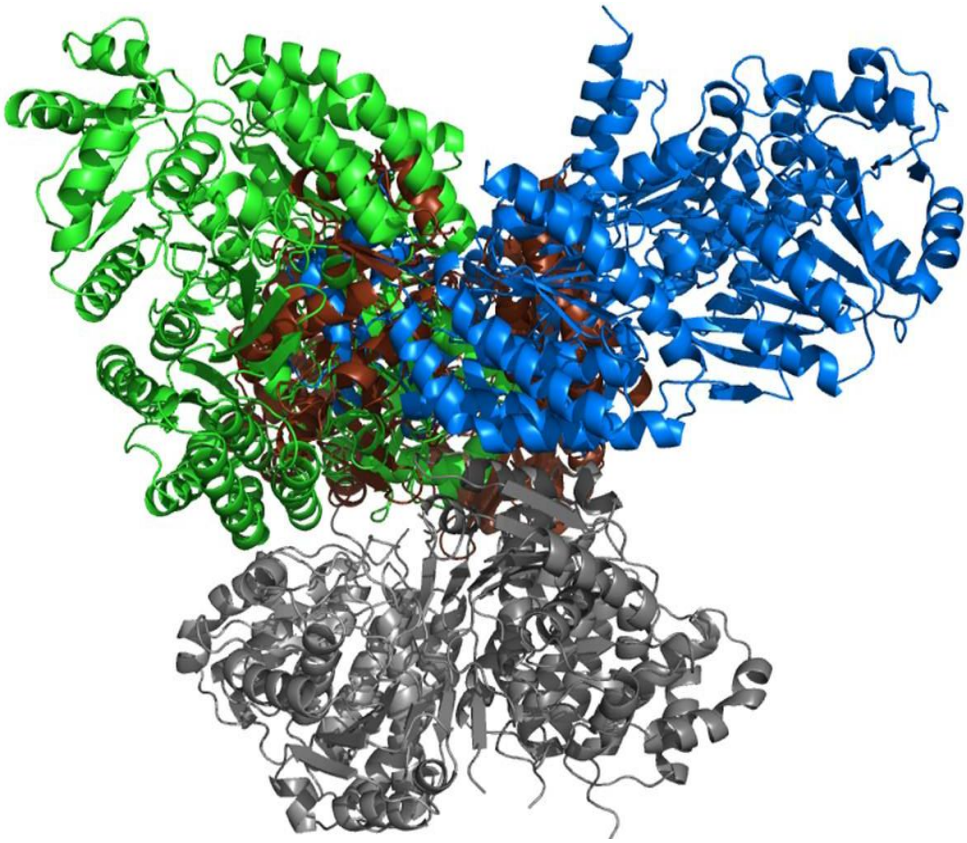


Figure 5-34: Cartoon overview of the predicted interactions between the modelled aldDH and Fragment 11 dimers. Grey = Fragment 11, brown = modelled aldDH top Hex result, blue = modelled aldDH top ClusPro result, and green = modelled aldDH top ZDOCK result.

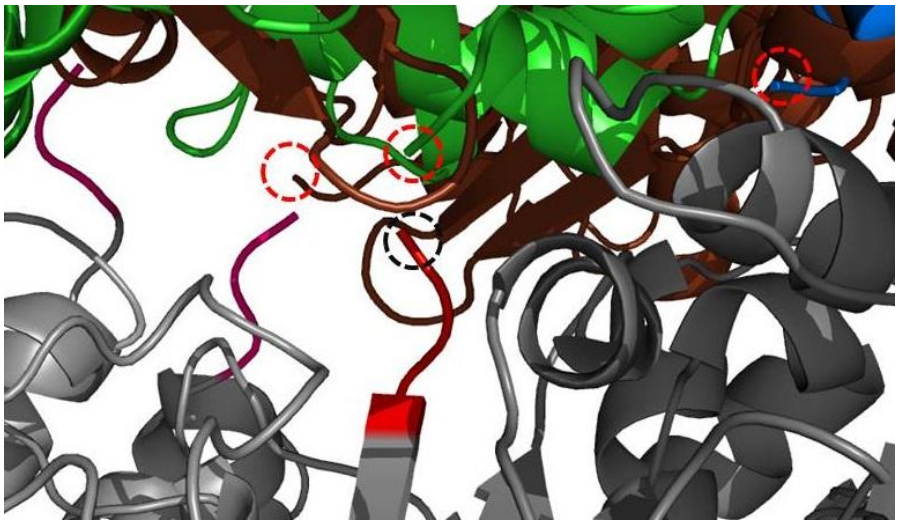


Figure 5-35: Cartoon diagram for the predicted interaction interface between the modelled aldDH and Fragment 11 dimers. Grey = Fragment 11, brown = modelled aldDH top Hex result, blue = modelled aldDH top ClusPro result, and green = modelled aldDH top ZDOCK result. Red dashed circles indicate the C-termini of modelled aldDH domains; the black dashed circle indicates the N-terminus of Fragment 11. Pink termini indicate the missing loop of Fragment 11.

PISA analysis was carried out on the top-rated model from each of the programmes used to evaluate the predicted interfaces between the two domains of ADHE.

Model generated with:	Interface area (Å^2)	Δ^iG (kcal/mol)	Δ^iG P-value	#HB	#SB
ZDOCK	2026.5	-16.7	0.452	13	0
Hex	1798.7	-1.9	0.623	9	6
ClusPro	1704.4	-6.2	0.99	0	21

Table 5-11: Summary of PISA analysis. Δ^iG = solvation free energy gain upon interface formation (-ve = hydrophobic interface). P-value = probability that the decrease in Δ^iG through the interface is not random, where P>0.5 = non-specific, and P<0.5 shows interface that may be considered interaction specific. HB = hydrogen bonds formed, and SB = salt bridges formed.

All the programmes independently predicted a similar interaction face between the aldDH and ADH domains as shown in Figure 5-35. These modelling results indicate that this face of the ADH domain of the protein is the most likely to be forming physiologically relevant interactions in the full ADHE protein with an aldDH domain.

There is no gap in the sequence between the modelled N-terminal aldDH domain and the C-terminal ADH domain; the gap between the two termini in the unlinked model would therefore be predicted to be short. The result of the ZDOCK programme positioned the aldDH C-terminus 7.4 Å away from the ADH N-terminus while the other two, although on the same face were more distant. This indicates that the modelled interaction by this programme may be physiologically relevant for the ADHE protein. PISA analysis suggested this model was also the most likely to be physiologically relevant although it should be noted that even this model had a relatively high P-value.

5.4.5.3 ZDOCK "ADHE" model interpretation

As the ZDOCK model gave the most promising docking result, further interpretation of the interactions between the modelled aldDH domain and the ADH domain of ADHE was based on this model. This model suggested that the dimerization interactions made by the aldDH and ADH domains of a single ADHE may be with different ADHE monomers, rather than within the same monomer (Figure 5-36).

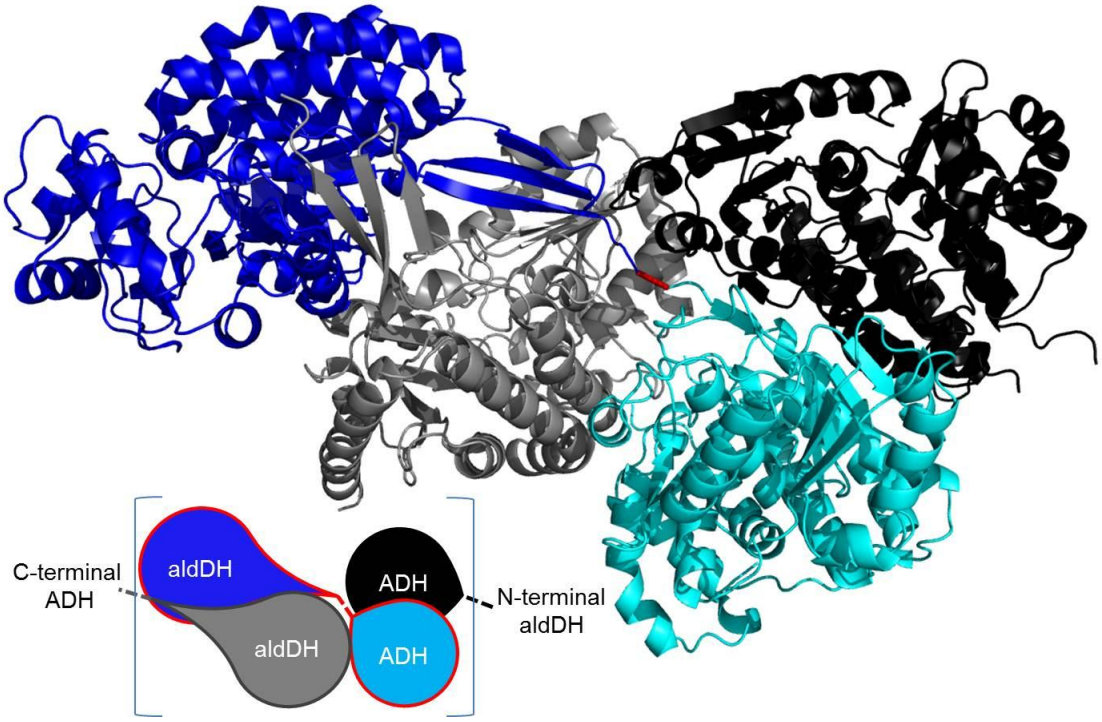


Figure 5-36: Cartoon overview diagram of the top result for the predicted interaction between modelled aldDH and Fragment 11 dimers using ZDOCK. Dark blue = modelled aldDH of ADHE 1, light blue = Fragment 11 of ADHE 1, grey = modelled aldDH monomer, and black = Fragment 11 monomer. The hypothesised link between the C-terminus of the aldDH and the N-terminus of Fragment 11 is shown in red. A schematic diagram in the same colour scheme is shown where one ADHE monomer is outlined in red.

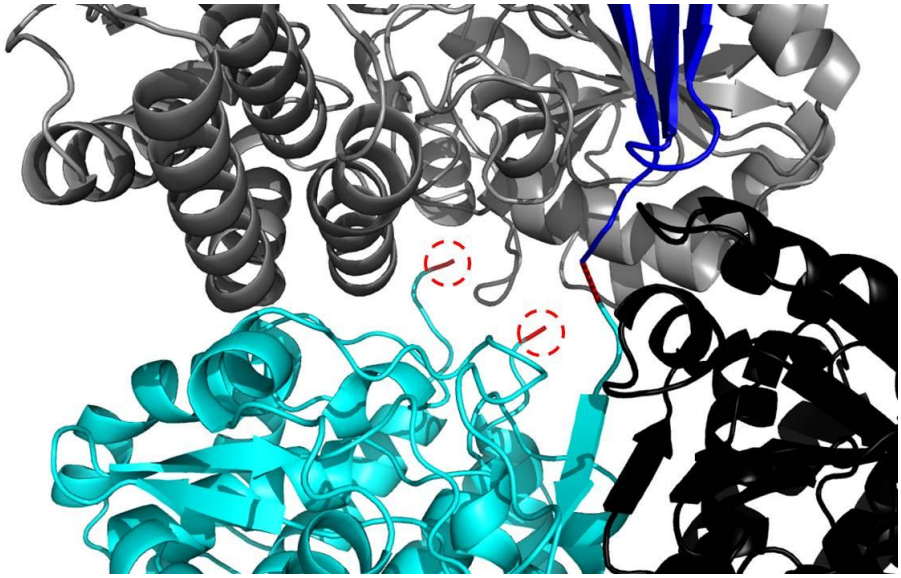


Figure 5-37: Cartoon diagram of the interface region for the top result for the predicted interaction between modelled aldDH and Fragment 11 dimers using ZDOCK. Dark blue = modelled aldDH of ADHE 1, light blue = Fragment 11 of ADHE 1, grey = modelled aldDH monomer, and black = Fragment 11 monomer. The hypothesised link between the C-terminus of the aldDH and the N-terminus of Fragment 11 is shown in red. Red circles indicate the termini of the truncated loop in the Fragment 11 structure.

The most mobile loop observed in the Fragment 11 crystal structure (loop F) did not appear to be directly involved in the modelled interaction with an aldDH domain. This loop may be sufficiently mobile to allow interaction *in vivo,* but this could not be predicted using the model derived here. The second most mobile loop (loop B) and the missing loop region in the Fragment 11 structure are both interacting in the model, suggesting these loops may be stabilised through interaction with an aldDH domain.

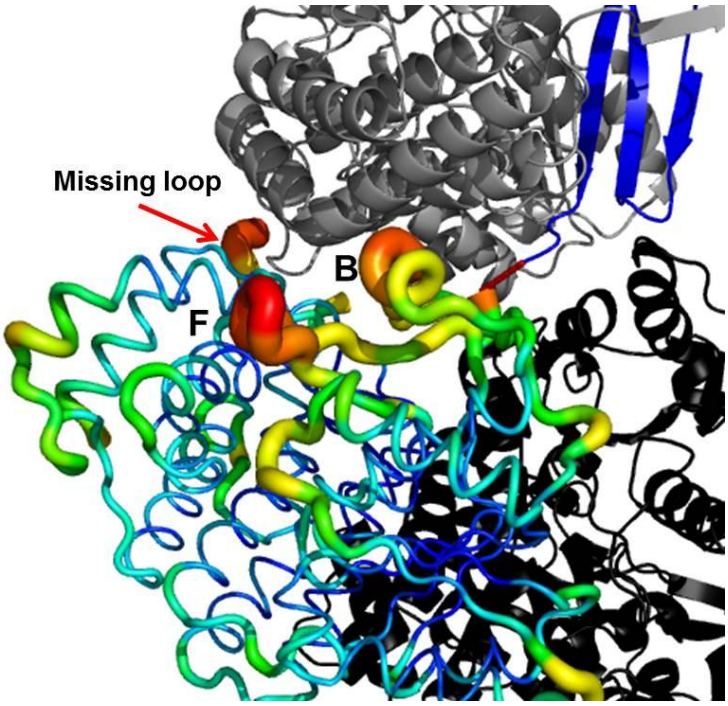


Figure 5-38: ZDOCK modelled interaction between Fragment 11 (visualised by temperature factor) and a modelled aldDH domain (grey in cartoon view). Wider and redder regions indicate increased mobility compared to thinner blue regions that indicate limited mobility. Loop labels correspond to those described in Figure 5-16. The aldDH domain predicted to be linked to the Fragment 11 monomer is shown in blue (cartoon view) with the termini linked by a red line; the second Fragment 11 monomer is shown in black (cartoon view).

The sequence of the missing loop in the Fragment 11 structure is KPKKFTAFPKYEYFK (the 8 missing residues in the structure are underlined). The whole loop is positively charged and enriched in aromatic residues. To identify possible stabilising residues at the predicted interface of the two domains, the solvent accessible areas for residues of the modelled aldDH dimer were calculated using AREAIMOL (CCP4i). Negatively-charged amino acids (Glu and Asp) and aromatic amino acids (Phe, His, Tyr and Trp) that had an accessible solvent area greater than 20 Å, were mapped onto the protein structure (range of accessible solvent areas observed ranged from 0-198 Å). As shown in Figure 5-39, several negative amino acids and aromatic residues are present close to the modelled interface between the

missing loop in Fragment 11, and the interacting aldDH monomer. This indicates the positively-charged, aromatic-enriched loop could feasibly be stabilised by residues in the interaction area proposed by the model.

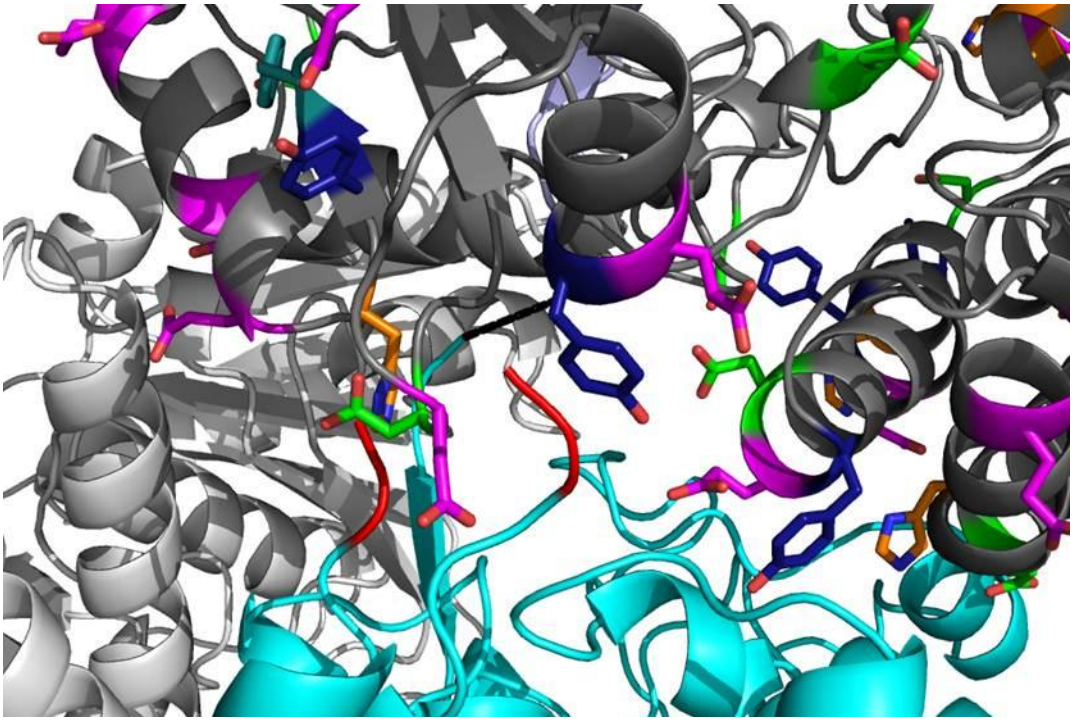


Figure 5-39: Cartoon diagram of the predicted interaction between the modelled aldDH and Fragment 11 dimers using ZDOCK, focussing on the missing loop region of Fragment 11. Faded blue = modelled aldDH of ADHE 1, light blue = Fragment 11 of ADHE 1, dark grey = modelled aldDH monomer and light grey = Fragment 11 monomer. The hypothesised link between the C-terminus of the ADHE aldDH domain and the N-terminus of the corresponding Fragment 11 is shown with a black line. The termini of the missing loop of Fragment 11 in ADHE 1 are shown in red. Residues potentially involved in loop stabilisation are shown; dark blue = Tyr, orange = His (N in blue), teal = Phe, purple = Glu and green = Asp.

It was not possible to infer the possibility of substrate channelling between the two domains of ADHE using the model produced here. The conserved Cys_{257} of the aldDH domain of ADHE discussed in the Introduction to this thesis was used as the predicted active site of the domain. Figure 5-40 shows the significant distance between the active sites of the two domains (approximately 40 Å). For channelling to occur it would be predicted that the active sites would be proximal to one another. It is possible that interactions between the two domains including the missing loop region, may allow channelling to occur. Conformational changes upon substrate binding may also affect the potential for channelling, although without a high resolution crystal structure of the ADHE protein this cannot be accurately determined.

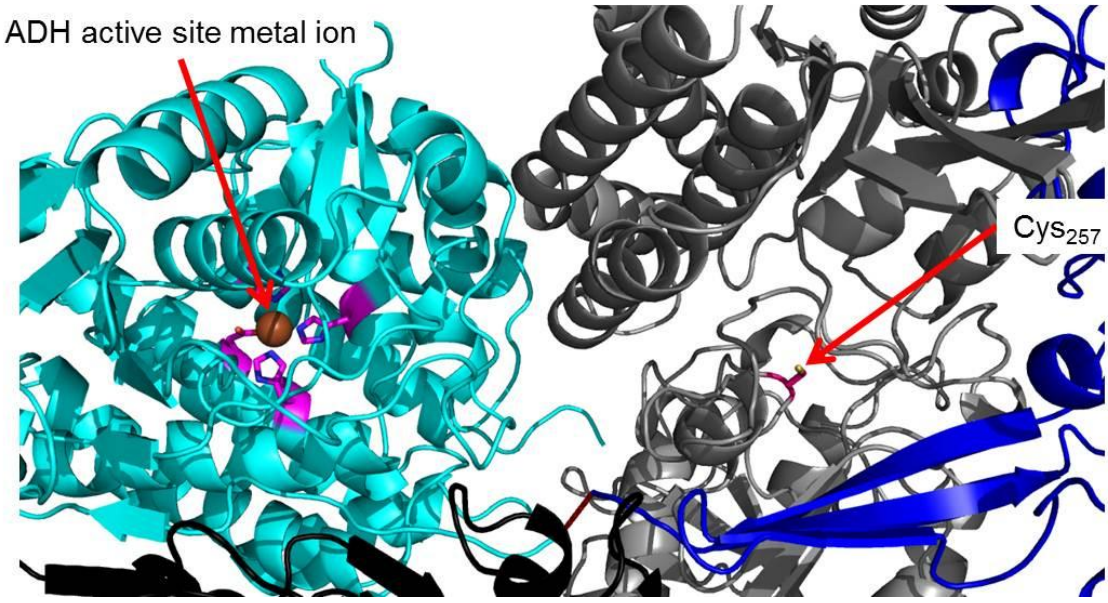


Figure 5-40: Cartoon overview diagram of the top result for the predicted interaction between modelled aldDH and Fragment 11 dimers using ZDOCK. Dark blue = modelled aldDH of ADHE 1, light blue = Fragment 11 of ADHE 1, grey = modelled aldDH monomer, and black = Fragment 11 monomer. The hypothesised link between the C-terminus of the aldDH and the N-terminus of Fragment 11 is shown in red. The predicted active site locations of the two domains are indicated with red arrows. The distance between the aldDH Cys_{257} and the active-site metal ion of the ADH domain is approximately 40Å.

When the interactions predicted between the two domains of ADHE shown in Figure 5-36 are extrapolated, a helical assembly of ADHE monomers can be observed (Figure 5-41 and Figure 5-42). Significant differences exist between this assembly and the spirosome structures observed by Kessler et al (1992) by electron microscopy. These authors report a left-handed helix, with 4 subunits per turn, whereas this model is right handed with 7 subunits per turn. Large multimeric assemblies of ADHE monomers were also observed during characterisation of the protein in Chapter 3. Variation in the interface angles generated by the model could cause the significant differences observed compared to the electron microscopy images. However, the potential for the modelled ADHE to form helical assemblies is an intriguing observation, although without a high-resolution structure of the ADHE protein, coupled with an electron density map of the spirosome assemblies, the models described remain purely speculative.

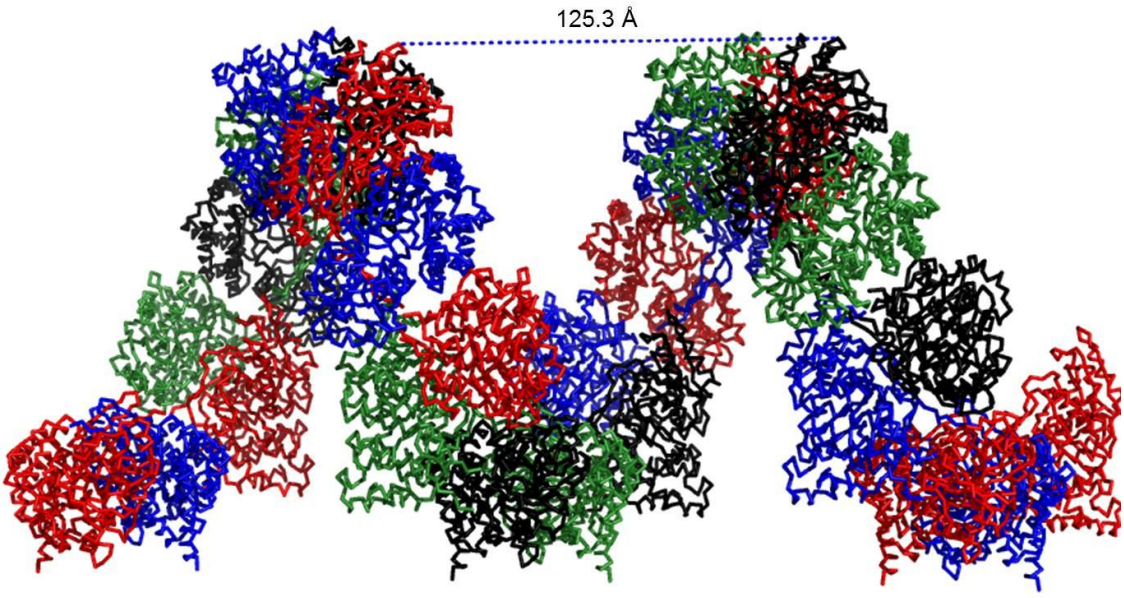


Figure 5-41: Cα trace of ADHE assembly based on the ZDOCK model in side-on view. Colours of the predicted ADHE monomers alternate through the Figure (right to left red-blue-black-green…). Approximately 7 monomers make up a whole turn, taking 125 Å to complete.

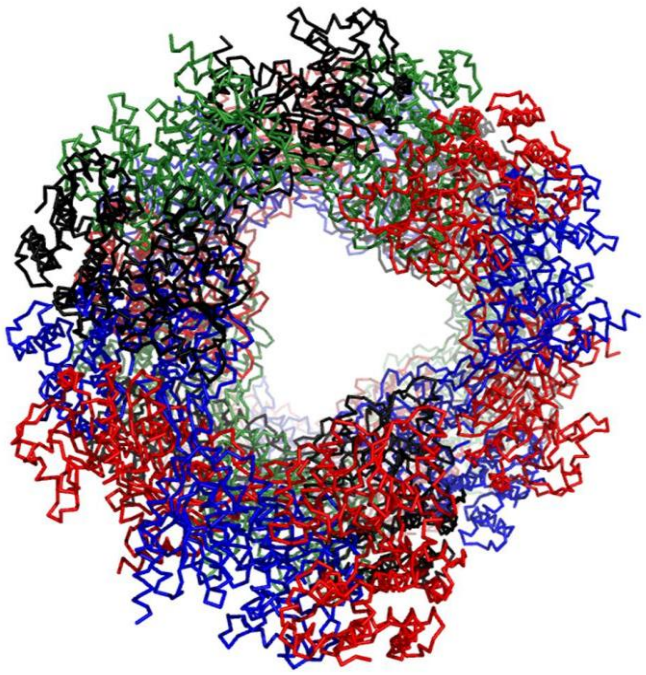


Figure 5-42: Cα trace of ADHE assembly based on the ZDOCK model in end-on view. Colours of the predicted ADHE monomers alternate through the Figure (right to left red-blue-black-green…).

5.5 Discussion

The ADH domain of the *Geobacillus thermoglucosidasius* ADHE protein has been resolved structurally using X-ray crystallography to 2.49 Å resolution. The protein has

been shown to consist of an NAD^+ binding domain (Rossmann fold), and an α-helical domain containing residues that are coordinating a metal ion. This metal ion is likely to be catalytic in nature due to its positioning at the interface between the two domains.

The metal ion identified in the structure of Fragment 11 is also present in similarly structured proteins in the PDB. The increased rate of catalysis observed for the ADH domain in Chapters 3 & 4 in the presence of divalent metal ions suggests this metal ion is likely to be catalytic in nature. As illustrated in Figure 5-43, the role of the divalent metal ion in the active site of the protein is probably to aid polarisation of the acetaldehyde carbonyl oxygen, allowing reduction by NADH to proceed. The identity of the metal ion may have an effect on the rate of catalysis due to differences in the strength of the polarisation of the carbonyl group. However, several different metal ions may be able to perform this role, the physiologically relevant metal ion present in the ADH domain of ADHE has not been unambiguously identified. Moreover, the intracellular conditions within *G. thermoglucosidasius* may influence the ion identity *in vivo*.

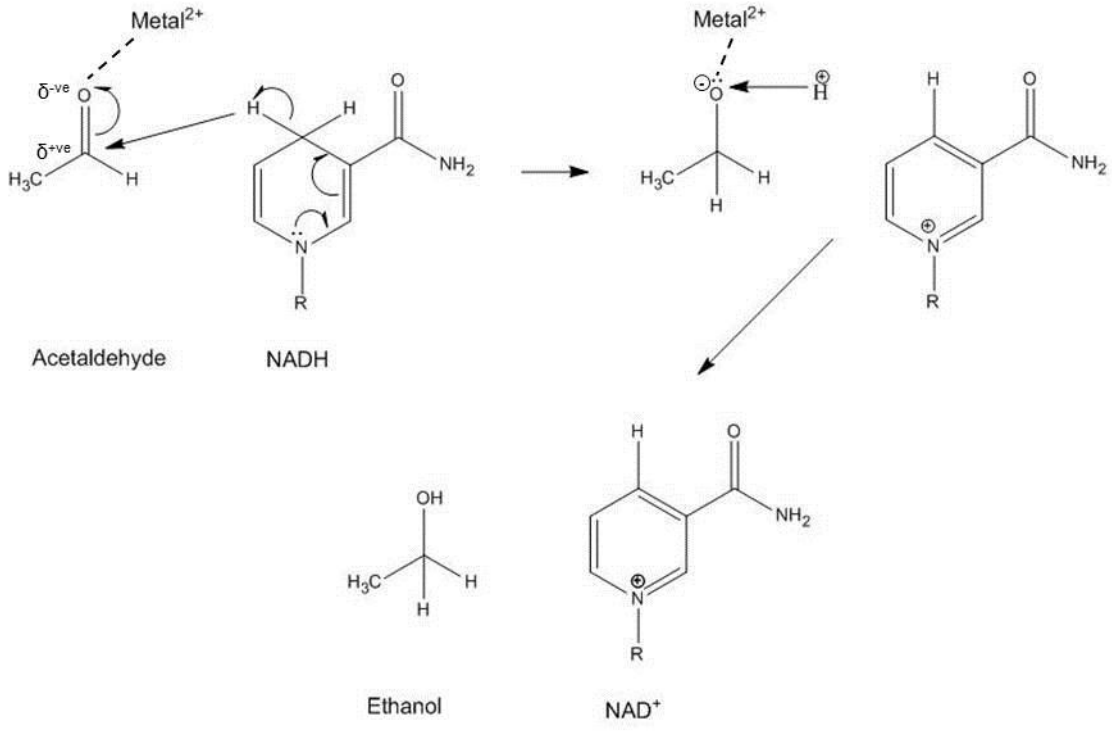


Figure 5-43: Reaction mechanism for the ADH catalysed reduction of acetaldehyde to form ethanol in the presence of a divalent metal ion.

Comparison of the Fragment 11 crystal structure to published proteins with similar folds in the PDB, showed that the overall protein structures were reasonably conserved (Table 5-7). Differences observed were mainly concentrated around the loop regions

of the proteins as shown in Figure 5-15 & Figure 5-16. Although significant variation in loop regions of proteins is common amongst homologous structures, the difference in length seen in some of the loops is considerable. In Fragment 11 the missing loop (loop C) is approximately 14 residues longer than the loops observed in 1O2D, 3BFJ and 1RRM. The loop sequence is KPKKFTAFPKYEYFK where missing residues in the structure are underlined. Several aromatic residues are present and the loop is positively charged. If the aldDH domain were to have a corresponding pocket enriched in negatively-charged and hydrophobic-residues in which this loop could reside, this may form part of a stabilising interaction between the two domains of the protein. Without the resolved aldDH domain of ADHE it is not possible to accurately determine if such a region exists. The loop is not close to the active site of the protein so does not appear to be directly involved in catalysis. *In silico* modelling of Fragment 11 with the modelled aldDH domain of ADHE, suggests that a suitable region may exist in the interacting aldDH domain of the protein that may act to stabilise this loop.

The unusually high temperature factors associated with several of the loop regions of the Fragment 11 protein, coupled with the limited thermostability observed during biochemical characterisation (Chapter 4), indicate that some stabilising interactions may exist between the two domains of ADHE. Exposed flexible loop regions within proteins can be susceptible to degradation and play a role in the instability of a protein at high temperatures (Nagi and Regan 1997). It is therefore common in the case of thermophilic proteins to limit the flexibility of loop regions to enhance stability. This may be through loop shortening (Russell et al. 1997), or by stabilisation of the loops through oligomerization (Vieille and Zeikus 2001). In the case of ADHE it is hypothesised that interactions with the N-terminal sister domain may be key to stabilising these mobile regions. It can be postulated that the aldDH domain may be orientated to interact with the more mobile side of the protein, thus stabilising the mobile loops. *In silico* modelling suggests that some of these mobile loop regions may be stabilised through interactions with an aldDH domain. Interestingly, this work suggests that such interactions may be inter-molecular, rather than intra-molecular, i.e. between different ADHE monomers. This may also help explain the spirosome assemblies that have been observed for ADHE proteins. Without a high resolution crystal structure of the ADHE protein, the hypothesised interactions between the aldDH and ADH domains of ADHE proteins cannot be validated.

Modelling of Fragment 11 with the $NADP^+$ molecule from the structurally similar 1O2D, provides some indication of residues likely to be interacting with $NAD(P)^+$ during

substrate binding. The actual conformation of the cofactor may differ significantly from the model. Comprehensive identification of key interactions involved in cofactor binding would therefore require a high-resolution crystal structure in the presence of $NAD(P)^+$. Although attempts were made to introduce NAD^+ into the crystals prior to collection of diffraction data, these attempts were unsuccessful. Co-crystallisation of Fragment 11 with NAD^+ may allow a high-resolution, cofactor-bound structure to be determined.

As discussed in the introduction to this Chapter, a recent publication by Brown et al (2011) described the effect of two point mutations (Pro_{704} to Leu_{704} and His_{734} to Arg_{734}) within the ADH domain of an ADHE from *Clostridium thermocellum.* These mutations were shown to convey an increased ethanol tolerance to this organism. The mutation identified as responsible is a histidine (His_{734}) to an arginine close to the active site of the protein (see Figure 5-44). This mutation was shown to convey a 25-fold reduction in ADH activity with respect to NADH. This was coincident with a 5-fold increase in specific activity in the presence of NADPH, as well as the increased ethanol tolerance of the organism.

It should be noted that in the experiments described by Brown et al (2011), the specific activity of the mutated protein with NADPH (0.12 µg/min/mg) remained significantly lower than the original specific activity of the un-mutated protein with NADH (2.7 µg/min/mg). This shift in cofactor specificity should be considered minor compared to the significant decrease in activity observed with NADH.

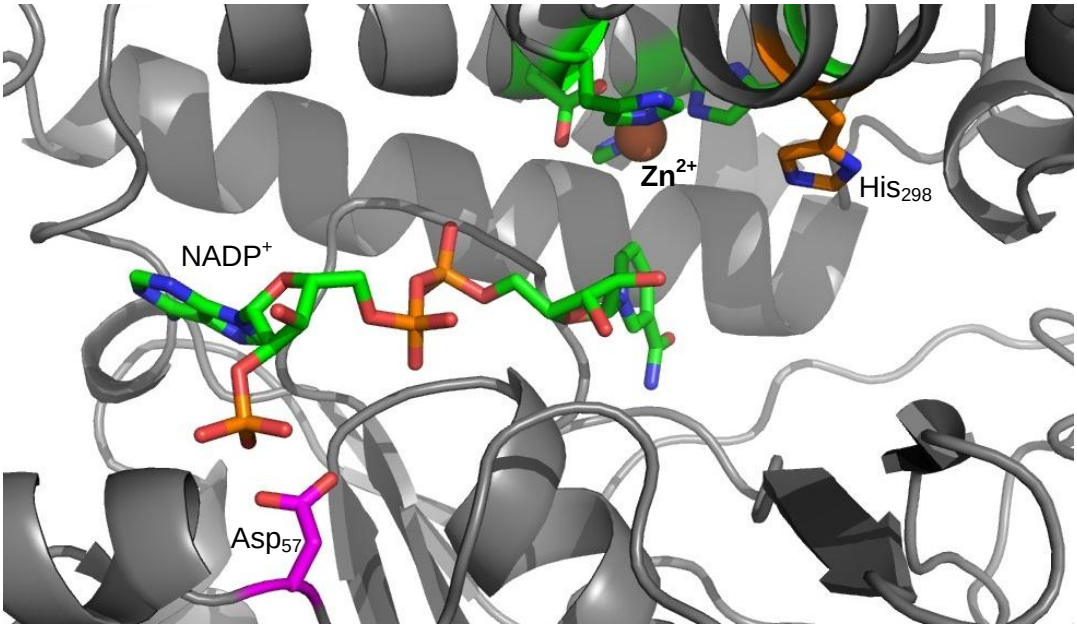


Figure 5-44: Cartoon of $NADP^+$ superimposed into the Fragment 11 crystal structure. Spirals = α-helices, arrows = β-strands, magenta residue = Asp_{57} & orange residue = His_{298} (equivalent to His_{734} discussed in Brown et al (2011). The superimposed $NADP^+$ and active-site metal ion are indicated.

The ADH domain of the *C. thermocellum* ADHE is 50% identical and 70% similar to Fragment 11. Conserved active site residues were identified through sequence alignments, this includes His_{298} of Fragment 11 which is equivalent to His_{734} in *C. thermocellum* (residue numbers differ significantly as Fragment 11 is the ADH domain only). Given the significant effects of mutations discussed in Brown et al (2011) on the host organism and the enzyme, the results will be interpreted in the light of the Fragment 11 crystal structure.

The significant decrease in activity observed may be caused by the introduction of a positive charge in the active site region (depending on the pH of the residue environment) possibly disrupting interactions with cofactor or the acetaldehyde substrate. A key interaction that may be disrupted is between His_{298} and Asp_{407}, which may have some influence over the conformation of the active site. The two residues are interacting in the Fragment 11 structure as shown in Figure 5-45. The Asp_{407} residue is also conserved in the sequence of the *C. thermocellum* ADHE (Asp_{844}), so the disruption of this interaction through mutation to an arginine may cause the changes in activity observed for the protein. The longer side-chain length may also cause disruption of the interaction between the active-site metal ion and the binding substrate, resulting in the significant decrease in activity.

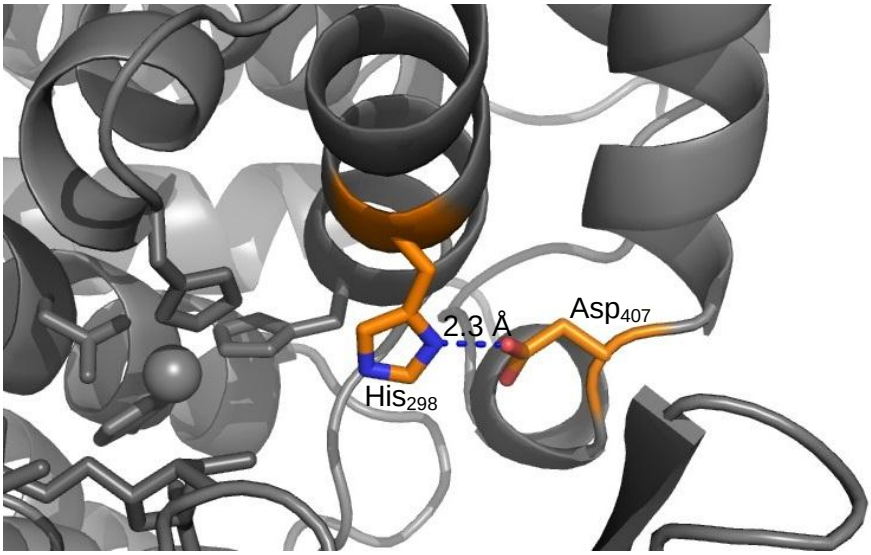


Figure 5-45: Cartoon of the Fragment 11 crystal structure around the His_{298} region. Spirals = α-helices, arrows = β-strands, orange residues = His_{298} (equivalent to His_{734} discussed in Brown et al (2011) and Asp_{407} (equivalent to Asp_{844} in the *C. thermocellum* ADHE). The distance between the interacting residues is indicated.

Lactate dehydrogenase is an example of an enzyme where cofactor switching between NADH and NADPH has been successfully carried out (Richter et al. 2011; Tomita et al. 2006). Richter et al. (2011) highlight a conserved Asp_{38} residue in the NADH specific *B. subtilis* LDH, and steric and electrostatic clashes with the additional 2' phosphate were suggested to prevent NADPH binding. This is a common feature that has previously been identified as potentially key in distinguishing NAD^+ and $NADP^+$ in enzyme active sites (Carugo and Argos 1997). Mutation of a proximal residue (Val_{39}) to a large positively charged residue (Arg), was shown to allow (and improve) catalysis in the presence of both NADH and NADPH. This mutation may have significantly changed the structural orientation of the originally-clashing loop region, allowing interactions with the additional phosphate of $NADP^+$ to form. Tomita et al (2006) showed for the LDH from *Thermus thermophilus,* that substituting a loop from an NADPH dependent malate dehydrogenase (removing the clashing Asp containing region), shifted the cofactor specificity of the enzyme. The mutations required to shift the cofactor specificities of these enzymes appear close to the point of interaction of the additional phosphate group of NADPH. However, the cofactor shifting mutations described by Brown et al (2011) are not close to the predicted site of interaction between the additional phosphate group of $NADP^+$ and the ADH domain of ADHE.

A key residue that may be controlling cofactor specificity in the Fragment 11 protein is Asp_{57}. This residue is also conserved in the *C. thermocellum* ADHE amino acid sequence (Asp_{494}). Activity of the enzyme with $NADP^+$ may be limited due to the steric

and electrostatic clash between the additional phosphate group and the Asp side chain. This residue is also likely to be a stabilising influence on the diol of the ribose group closest to the adenine ring of the NAD^+ molecule. Given the distance between His_{298} and the additional phosphate group (18.4 Å), it is difficult to resolve how the increased activity with respect to NADPH is caused by this mutation in the *C. thermocellum* ADHE. A high-resolution crystal structure of the ADH domain of this mutant protein in the presence of $NADP^+$ and NAD^+ could provide significant insight as to the effect of these mutations on catalysis. Interestingly, another publication by the same group suggests that another ethanol tolerant strain of *C. thermocellum* showed a mutation in this Asp_{494} residue, although biochemical evaluation of this mutant version was not described (Shao et al. 2011).

The effect of the mutations on the host organism's ethanol tolerance is intriguing, as conventional understanding of ethanol tolerance is influenced by cell membrane properties (Liu and Qureshi 2009). Brown et al (2011) suggest a link between electron flow and membrane properties that may be influenced by a shift in cofactor specificity of metabolic enzymes. This hypothesis was not further investigated as part of this project.

An appealing future direction may be to monitor the effect of targeted mutagenesis of the ADH domain of ADHE, in terms of ethanol tolerance and cofactor specificity. The structure determined for Fragment 11 would provide valuable information to determine specific target residues for such mutagenesis. As well as evaluating the effect of mutating His_{298}, a potential target for cofactor switching could be Asp_{57}, changing to a smaller uncharged residue such as serine or glycine or alternatively, the introduction of a positively charged residue such as arginine or lysine, may also facilitate interaction with the additional phosphate group of NADPH.

In conclusion, the structure of the ADH domain of the TM242 ADHE protein has been successfully resolved during this section of the project. The dimeric protein appears to require stabilisation by the aldDH domain of the protein for optimal thermostability, due to the presence of highly mobile loop regions within the structure. Simulated docking experiments have suggested that these interactions may be between different ADHE monomers rather than between the two domains of an ADHE protein. Modelling of an $NADP^+$ cofactor into the structure has indicated some key residues likely to be involved in cofactor binding. The presence of a divalent metal ion has been confirmed at the active site of the protein. Analysis of the structure indicates that the loss in activity

observed for a mutant of the similar ADHE protein of *C. thermocellum* (Brown et al. 2011), may be due to disruption of regions of the protein close to the active site. The basis of the observed change in cofactor preference could not be determined on the basis of this structure without bound co-factor.

6 IDENTIFICATION & CHARACTERISATION OF OTHER FUNCTIONAL ACETYLATING ALDEHYDE DEHYDROGENASES FROM *GEOBACILLUS THERMOGLUCOSIDASIUS*

6.1 Introduction

Efficient metabolism of acetyl-CoA to acetaldehyde is a key factor in ethanol production in TM242. However, within cell extracts the activity detected for the aldDH domain of ADHE was consistently lower than that observed for PAT (Hills, C 2011 unpublished work; TMO Renewables 2009 personal communication). Within fermentative metabolism, competition between these two enzymes for the acetyl-CoA was suggested as a potential "bottle-neck" in ethanol production. Characterisation of the ADHE protein from TM242 (Chapter 3) showed the aldDH domain to be relatively unstable compared to the ADH domain. This domain also appeared to be subject to substrate inhibition, limiting the maximum rate of catalysis. Assuming the relatively high K_m for the ADH domain with respect to acetaldehyde may be overcome through substrate channelling, the aldDH activity was identified as a potentially limiting step in ethanol production. *In vitro* evidence suggests that there may be an opportunity to optimise the metabolism of acetyl-CoA through the manipulation of the aldDH activity within the cell.

Attempts were made to resolve the aldDH domain of ADHE independently of ADH both structurally and kinetically (Chapters 4 & 5). Although soluble protein fragments were produced, no activity was detected and the most soluble of these fragments did not crystallise under the conditions tested. Efforts therefore focussed on identifying suitable independent aldDH proteins from the TM242 genome. Such proteins could provide a "metabolic boost" in terms of conversion of acetyl-CoA to ethanol, or could be used to replace the aldDH domain of ADHE in an attempt to produce a more catalytically active/stable protein fusion. Due to regulations regarding genetically modified organisms, it is preferential in a production strain for process applications to use genes already present in the organism and manipulate their expression/activity, rather than introducing genes from another organism (TMO Renewables 2009 personal communication).

This chapter describes the cloning of two acetylating aldDH-encoding genes identified in the TM242 genome, and characterisation of the recombinantly expressed enzymes.

6.2 Materials and methods

6.3 Gene identification

To identify possible aldDH protein coding genes in the TM242 genome, a protein BLAST search (of hypothesised gene products) was carried out based on the protein sequence of the ADHE protein Fragment 1 (predicted aldDH domain) using the ERGO (Integrated Genomics) tool.

6.3.1 AldDH protein cloning

TM242 genomic DNA (50 ng/µl) prepared by TMO Renewables was used as the DNA template for PCR-amplification of the genes described here. These were carried out according to the method described in Section 2.3.2. The T_m of the primers used was higher than for amplification from plasmid DNA so an annealing temperature of 77°C was used for these reactions. The DNA primer combinations used to amplify the genes are shown in Table 6-1. Primer DNA sequences are reported in Appendix 1. Primers incorporated additional bases where required to keep the desired tags in-frame.

Protein (expression vector)	Forward primer	Reverse primer	Restriction sites
EutE (pUCG18-pLDH)	eutEpUCG18F	eutEpUCG18R	XbaI/SacI
EutE (pET45b)	eutEpET45bF	eutEpET45bR	KpnI/XhoI
EutE (pLM303)	eutE pLM303	eutEpET45bR	KpnI/XhoI
AcAldDH (pUCG18-pLDH)	GB ACaldDH Fwd1	GB ACaldDH Rev1	XbaI/SacI
AcAldDH (pET28a)	ACaldDH pet28 F1	ACaldDH pet28 rev1	NheI/XhoI

Table 6-1: PCR primers used for the amplification of genes encoding acetylating aldDH proteins.

The PCR products were A-tailed, ligated into the pGEM®-T easy vector, ethanol precipitated and transformed into JM109 cells for blue/white screening. Successful cloning of the protein-coding genes was confirmed by DNA sequencing. The genes in pGEM®-T easy were then digested with the appropriate restriction enzymes, gel purified and ligated into the previously-digested protein-expression vectors. Ligations were transformed into JM109 cells that were then subjected to PCR colony screening. Positive plasmids from the PCR screen were screened by restriction digestion with the required enzymes, prior to positive constructs being sent for DNA sequencing with vector specific primers.

Protein expression vectors were transformed into the desired protein expression strains as detailed previously.

6.3.2 Maltose binding protein tag

During experiments with the EutE protein, a maltose binding protein (MBP) solubility tag was incorporated at the N-terminus of the protein using the pLM303 vector. This is a non-commercially available vector produced by the Vanderbilt centre for structural biology (Nashville TN, USA). This vector also incorporates a poly-histidine tag at the N-terminus of the MBP allowing the MBP-protein fusion to be purified by metal affinity chromatography. The MBP can also act as a purification tag through its interaction with amylose resin.

Amylose columns use the affinity of the maltose binding protein tag for maltose (amylose), to separate the proteins of interest from other proteins produced by the expression strain. The column was prepared according to the manufacturer's instructions using amylose-agarose resin (NEB). The column was washed with distilled water and equilibrated with HIS-BIND buffer prior to loading of the soluble protein fraction onto the column. The various fractions were collected in separate tubes and stored on ice.

The initial fraction eluted was reloaded to ensure that all the tagged protein became bound to the column. Following this, the column was washed with 20 column volumes of HIS-BIND buffer to elute any unbound material. An increasing concentration of maltose (in HIS-BIND buffer) was used to elute bound proteins stepwise in order of affinity (5 ml elution for each step). This protocol was adjusted where required to maximise the purity of the protein obtained. Following collection of the fractions, samples were taken for SDS-PAGE to confirm the presence and purity of the protein product obtained.

A "PreScission" protease site (GE) between the MBP and the target protein allowed cleavage of the solubility tag post-purification.

6.3.3 Enzyme assays

Unless otherwise stated, aldDH NADH assays (Section 2.13.1.2) were used to measure enzyme activity. Assays of the AcAldDH protein showed rapid curvature indicative of product inhibition (discussed in the Results section of this chapter); initial rates for this enzyme were therefore taken over the first 10-15 seconds of the assay.
Assays to determine the AcAldDH kinetic parameters in the reverse direction were carried out. These assays used citrate synthase as a coupling enzyme to rapidly

regenerate the CoA-SH, allowing accurate kinetic parameters to be determined (illustrated in Figure 6-1).

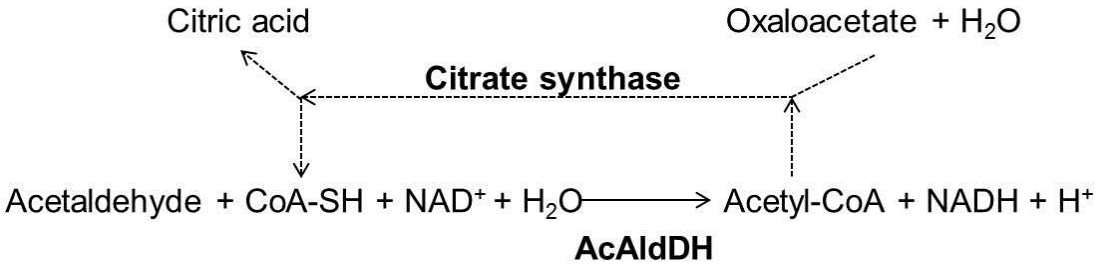


Figure 6-1: Overview diagram of the coupled AcAldDH reverse direction assay.

Chemical	Stock conc^n (mM)	Volume added (µl)	Assay conc^n (mM)
Citric acid buffer pH 6.0 0.1 mM zinc acetate	50	864*	~50
CoA-SH	6.5	50	0.33
Acetaldehyde	4000	5	20
NAD^+	10	50	0.5
Enzyme / cell extract		10	
Oxaloacetate	10	20	0.2
Citrate synthase	8.7mg/ml (≥100 U mg^{-1})	1	0.87 mg

Table 6-2: Reverse direction AcAldDH with citrate synthase assay conditions.

- *Buffer volume adjusted as required where volume of CoA-SH varied for assays
- Buffer pre-incubated at 60°C
- Substrates stored on ice
- Reaction started with protein sample addition

6.4 Results

6.4.1 Gene identification

Two additional aldDH coding genes were identified through the BLAST search of the TM242 genome. These were both annotated as belonging to EC 1.2.1.10 (acetylating aldehyde dehydrogenases).

The first protein, known as EutE, was an aldehyde oxidoreductase protein (25% identity, 43% similarity, 21% gaps to Fragment 1). This protein was annotated as a EutE protein, the name for which is derived from its involvement in ethanolamine metabolism in other organisms (Garsin 2010). The second, known as AcAldDH, was annotated as an acetylating aldehyde dehydrogenase (43% identity, 59% similarity,

12% gaps to Fragment 1). Both these proteins were selected for expression trials and characterisation.

The annotated start codon of the *eutE* gene was TTG; this was changed to the stronger start codon ATG for protein expression.

6.4.2 Gene cloning

The sizes of PCR-amplified DNA fragments for the various expression vectors were determined as shown in Figure 6-2; DNA bands of the expected size were excised and gel purified. The fragments were A-tailed and ligated into the pGEM®-T easy vector overnight.

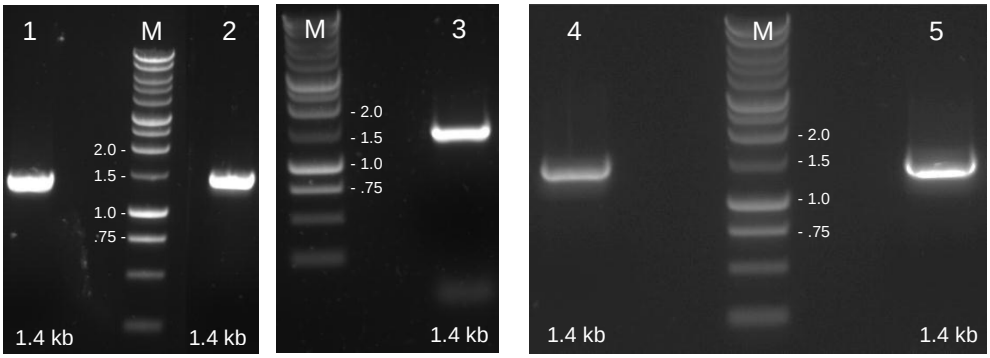


Figure 6-2: Agarose gel electrophoresis of PCR products. 1 = *eutE* for pET45, 2 = *eutE* for pUCG18-pLDH, 3 = *eutE* for pLM303, 4 = *AcAldDH* for pUCG18-pLDH, 5 = *AcAldDH* for pET28a. M = DNA markers with sizes given in kb. The predicted sizes of DNA fragments are shown below the appropriate lane.

Once the genes were confirmed to be successfully cloned into pGEM®-T easy, the DNA fragments were excised using the appropriate restriction enzymes, prior to ligation into the previously-digested and SAP-treated expression vectors. Ligations were ethanol precipitated prior to transformation into JM109. Transformants were screened using the antibiotic resistance markers for *E. coli* present in the vectors. Representative colonies for each were picked for overnight cultures prior to plasmid preparation. Initially the vectors were screened using restriction digestion. Positive results were confirmed by DNA sequencing using vector-specific primers.

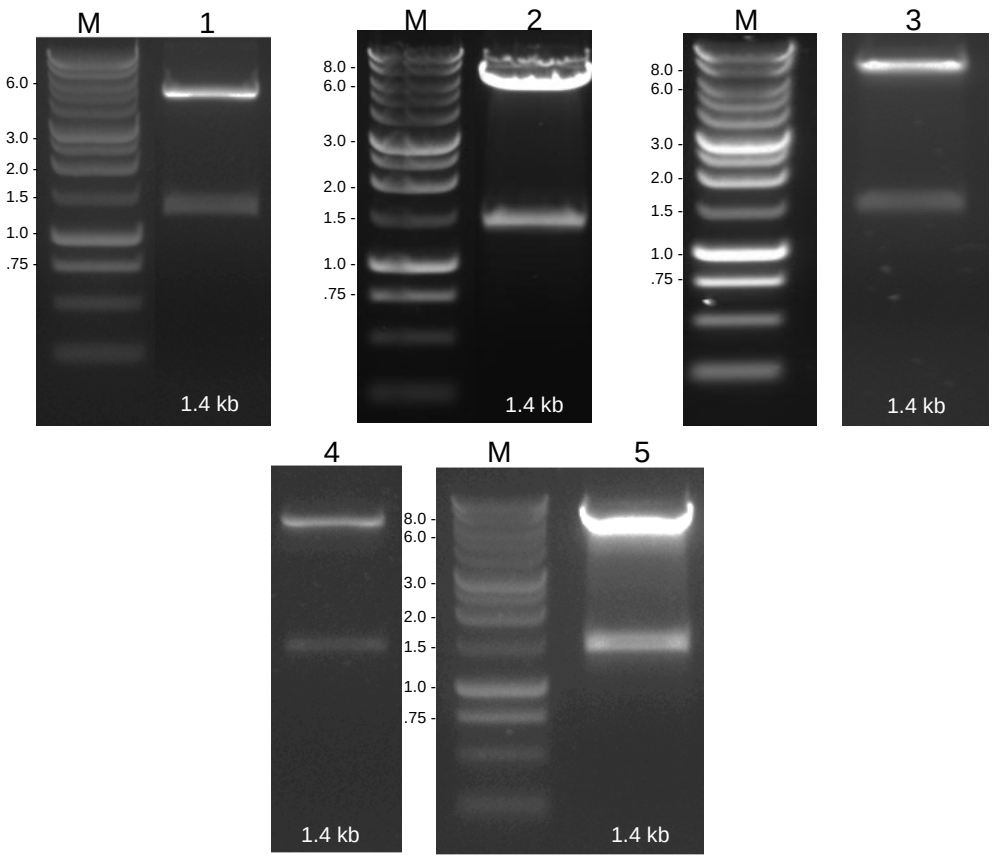


Figure 6-3: Agarose gel electrophoresis of expression vector restriction digests. 1 = *eutE* in pET45b, 2 = *eutE* in pUCG18-pLDH, 3 = *eutE* in pLM303, 4 = *AcAldDH* in pUCG18-pLDH, 5 = *AcAldDH* in pET28a. M = DNA markers with sizes given in kb. The predicted sizes of DNA fragments are shown below the appropriate lane.

6.4.3 Protein expression trials and protein purification

The purified plasmid constructs required were transformed into protein expression strains BL21 (DE3), Rosetta© (DE3) and Arctic Express© (DE3) *E. coli* strains.

6.4.3.1 EutE

The majority of the His-tagged EutE protein produced appeared to be in the insoluble fraction of the cell extracts in the *E. coli* protein expression strains. Expression of the untagged EutE was also tested in TM242 and TM400 grown both aerobically and anaerobically (produced as in Chapter 4), but the protein was still insoluble. A representative gel of the BL21 expression trial is shown in Figure 6-4 (A), where the insoluble protein consistently ran at a higher M_r than predicted for the protein. Low levels of aldDH activity were detected in the soluble cell extracts, suggesting that if the protein was folded correctly it may be active. It was decided to use an MBP solubility tag (introduced using the pLM303 vector) in an effort to induce soluble protein

expression (Kapust and Waugh 1999). The MBP tag successfully induced soluble expression of the EutE protein.

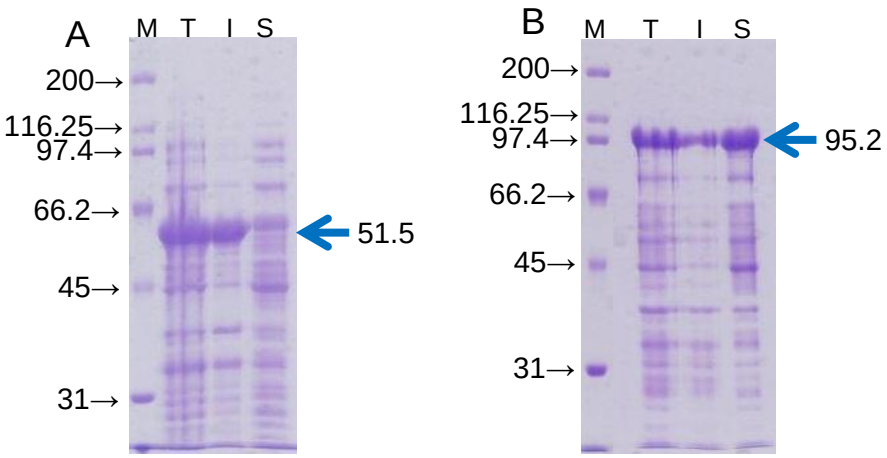


Figure 6-4: Representative SDS-PAGE analysis of BL21 protein expression trials. A = EutE (pET45b) and B = MBP-EutE (pLM303). M = markers (M_r/1000), T = total, I = insoluble & S = soluble (BL21 4.5 h expression shown). Predicted His-tagged-protein bands are highlighted with a blue arrow, and their expected M_r/1000 values shown.

The MBP-EutE fusion protein expressed in BL21 (DE3) (4.5 h induction) was used for subsequent experiments (Figure 6-4 B). The protein was initially purified from soluble cell extracts using metal affinity chromatography. Pooled fractions with the highest aldDH activity were then loaded onto the amylose column. A representative SDS-PAGE gel for this purification is shown in Figure 6-5.

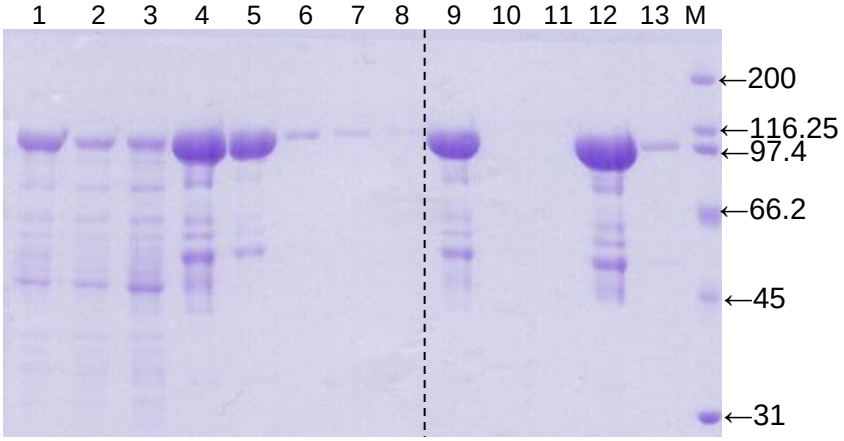


Figure 6-5: SDS-PAGE gel of MBP-EutE purification. 1-8 = Metal affinity samples, 1 = soluble, 2 = flow-through, 3 = 0%, 4 = 5%, 5 = 10%, 6 = 20%, 7 = 30% & 8 = 60% (% = % HIS-ELUTE buffer). 9-13 = amylose purification samples, 9 = load, 10 & 11 = 0 mM maltose, 12 = 10 mM maltose, 13 = 20 mM maltose. M = markers (M_r/1000).

The specific activity observed (using aldDH NADH assays) for the purified MBP-EutE fusion protein (sample 12 in Figure 6-5) was 3.2 U mg^{-1}. The 10 mM maltose elution sample was dialysed into HIS-BIND buffer for 6 h at 4°C and the MBP tag was then

successfully cleaved using the "PreScission" protease (GE) according to the manufacturer's instructions. The cleaved MBP protein was unable to be resolved from EutE through amylose or metal affinity purifications. Due to the issues with protein solubility and the relatively low observed specific activity for EutE, subsequent experimental efforts focussed on the AcAldDH protein.

6.4.3.2 AcAldDH

The AcAldDH protein was expressed in the soluble cell extract of all the protein expression strains tested. The BL21 (DE3) strain was selected for expression of AcAldDH using 4.5 h protein expression (Figure 6-6 A). The AcAldDH protein was purified using metal affinity chromatography as shown in Figure 6-6 B.

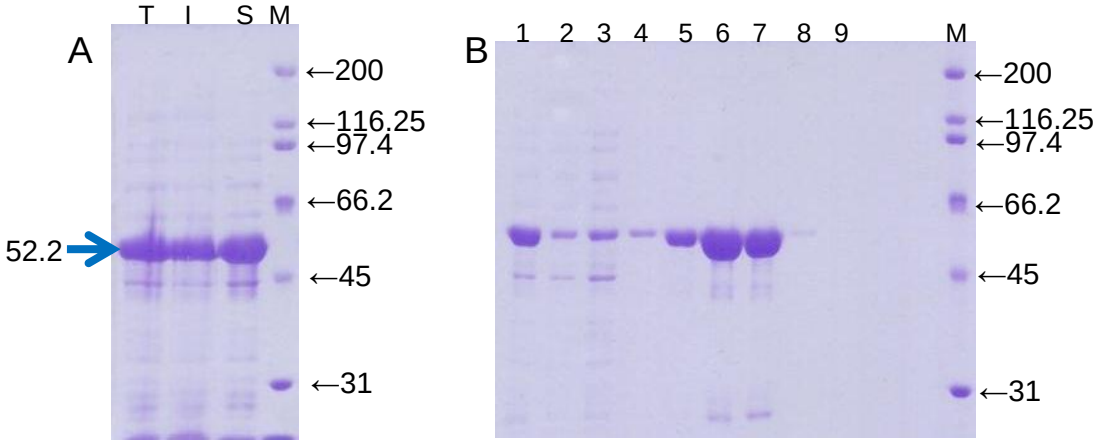


Figure 6-6: A = SDS-PAGE analysis of AcAldDH BL21 protein expression trial. M = markers (M_r/1000), T = total, I = insoluble & S = soluble (BL21 4.5 h expression shown). The predicted His-tagged-protein band is highlighted with a blue arrow, and the M_r/1000 of the expected protein is shown. B = SDS-PAGE gel of an AcAldDH metal affinity purification. 1 = soluble, 2 = flow-through, 3 = 0%, 4 = 2.5%, 5 = 5%, 6 = 10%, 7 = 20%, 8 = 30%, & 9 = 40% (% = % HIS-ELUTE buffer).

The 10 & 20% fractions from the metal affinity purification showed the highest aldDH activity. These samples were pooled prior to dialysis into 50 mM Tris-HCl pH 8.0, 150 mM NaCl after a 2 h dialysis against 50 mM Tris-HCl pH 8.0, 150 mM NaCl and 75 mM imidazole. This method of purification was used for all the AcAldDH protein samples used during this project.

6.4.4 Kinetic characterisation of the AcAldDH protein

AldDH NADH assays (NADH depletion) of the AcAldDH protein showed rapid curvature and the reaction did not go to completion; an example trace for this assay is shown in Figure 6-7. Initial rates for this enzyme were therefore taken over the initial 10-15 seconds where the reaction was linear in nature.

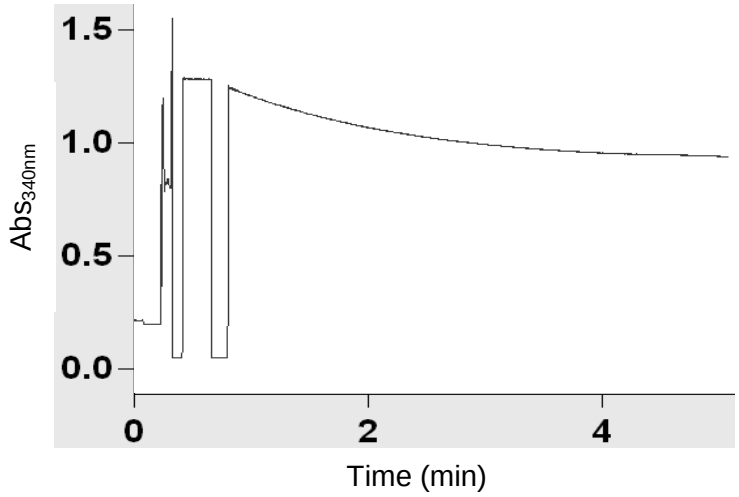


Figure 6-7: An example AcAldDH enzyme reaction trace (Abs_{340nm} vs time (min)).

Assays were undertaken to determine kinetic parameters for this enzyme under these conditions.

Activity (substrate)	K_m (mM)	Standard Error	V_{max} (U mg^{-1})	Standard Error
AcAldDH: (acetyl-CoA)	0.024	+/- 0.001	37.5	+/- 4.0
AcAldDH: (NADH)	0.042	+/- 0.002	34.5	+/- 3.8

Table 6-3: Summary of kinetic parameters determined for the AcAldDH protein. V_{max} reported was adjusted for the degree of saturation with the fixed substrate.

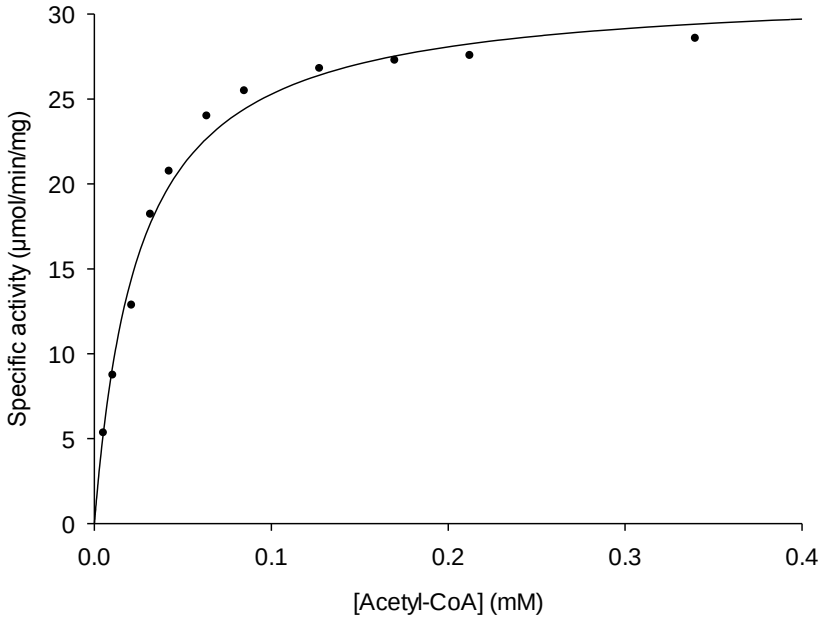


Figure 6-8: Michaelis-Menten plot of aldDH enzymic activity (U mg^{-1} of protein) against concentration of acetyl-CoA (mM) at a fixed concentration of NADH (0.21 mM) for the purified AcAldDH enzyme.

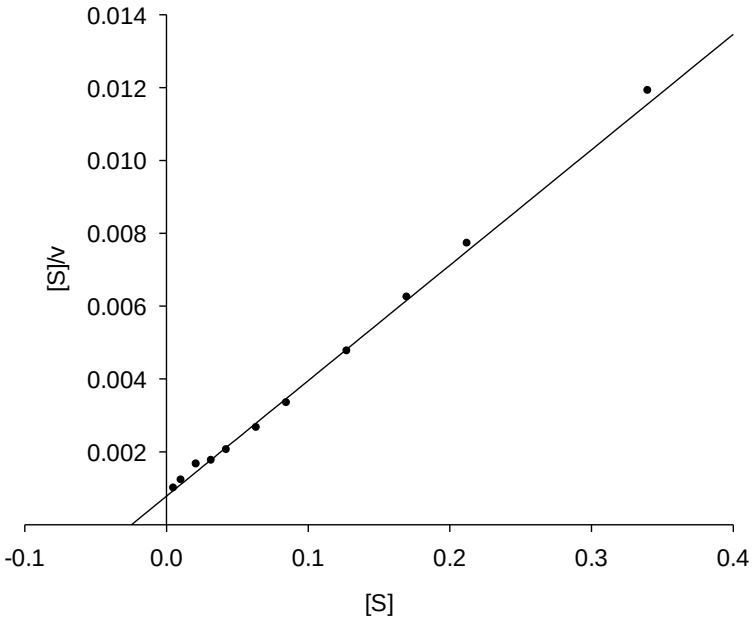


Figure 6-9: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of aldDH activity (U mg^{-1}) with respect to concentration of acetyl-CoA (mM).

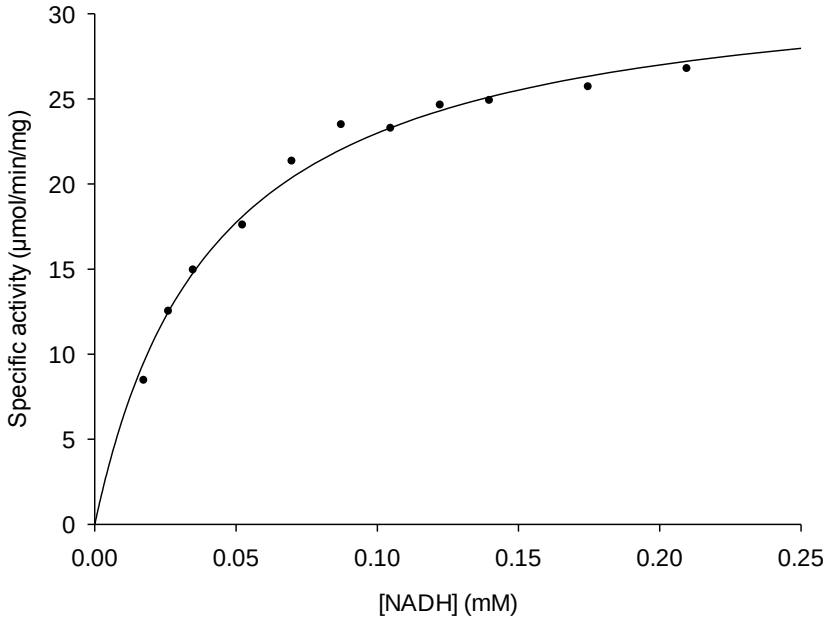


Figure 6-10: Michaelis-Menten plot of aldDH enzymic activity (U mg^{-1} of protein) against concentration of NADH (mM) at a fixed concentration of acetyl-CoA (0.34 mM) for the purified AcAldDH enzyme.

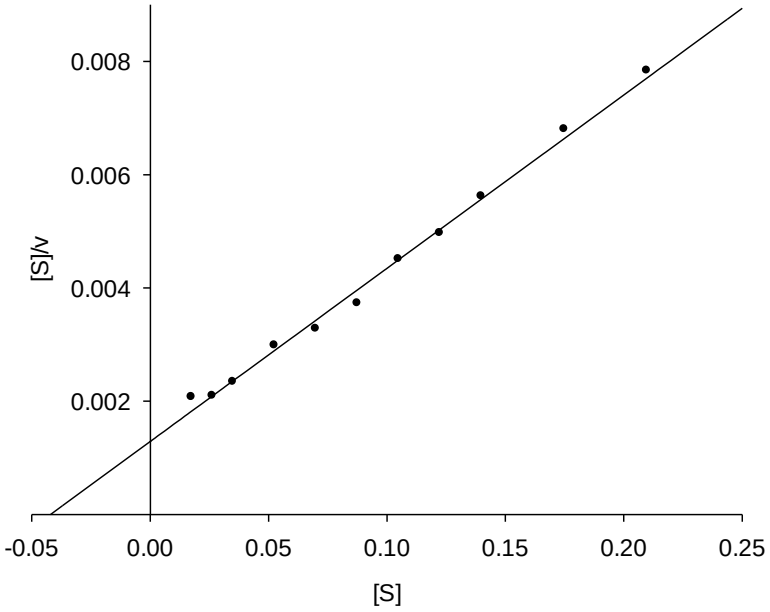


Figure 6-11: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of aldDH activity (U mg^{-1}) with respect to concentration of NADH (mM).

The rapid curvature of the progress curves observed during the assay of this enzyme suggested strong product inhibition. Assays were carried out to determine the nature of this inhibition.

6.4.5 Investigation into product inhibition of AcAldDH domain

Assays to determine which of the reaction products was causing the inhibition were carried out. The AcAldDH protein was assayed in the presence of different concentrations of each of the reaction products. Approximately 0.08 mM CoA-SH and NAD^+ had accumulated by the point of curvature of the progress curves; reaction products in this concentration range were therefore used for these assays. The specific activities observed in these assays appear higher than those observed in Section 6.4.4. The reason for this discrepancy is unknown; differences in protein purity between enzyme preparations is a possible explanation.

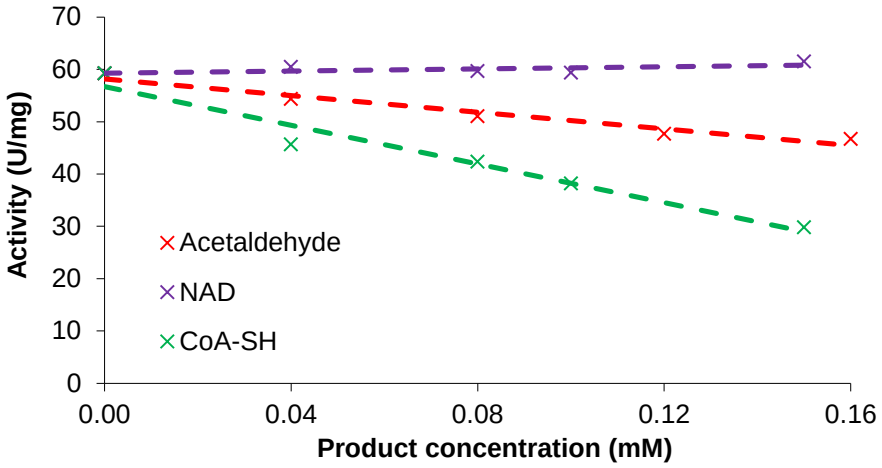


Figure 6-12: Initial rates of AcAldDH measured in the presence of reaction products.

CoA-SH and acetaldehyde were both shown to inhibit the forward reaction. As the effect of CoA-SH appeared more significant, this product was used as the inhibitor for subsequent experiments.

Kinetic parameters for the AcAldDH protein with respect to acetyl-CoA were re-determined in the presence and absence of 0.13 mM CoA-SH. The effect of this product on the kinetic parameters allowed the type of inhibition to be determined.

Activity	K_m (mM)	Standard Error	V_{max} (U mg^{-1})	Standard Error
AcAldDH: (0 mM CoA-SH)	0.032	+/- 0.001	57.7	+/- 0.55
AcAldDH: (0.13 mM CoA-SH)	0.054	+/- 0.001	36.7	+/- 0.21

Table 6-4: Comparison of kinetic parameters for the AcAldDH protein with respect to acetyl-CoA concentration, determined with and without 0.13 mM CoA-SH (a fixed concentration of 0.18 mM NADH was used in these assays).

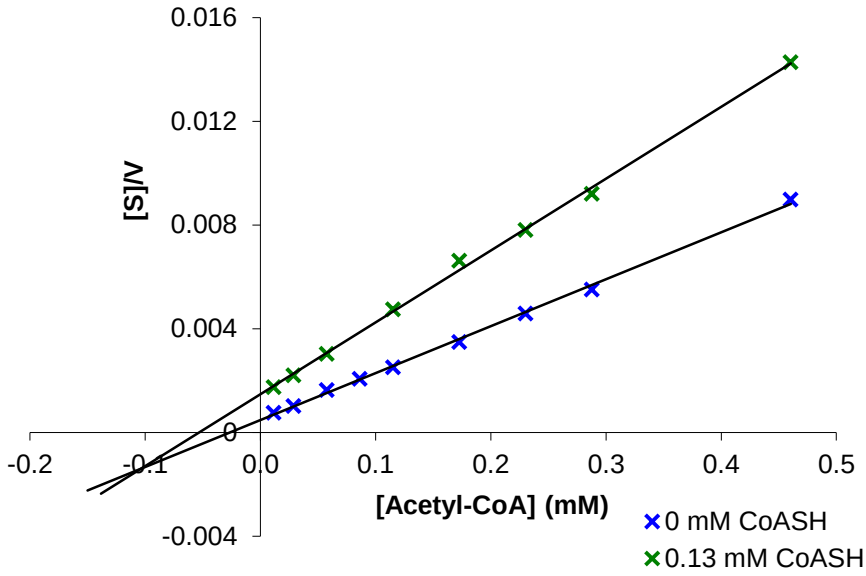


Figure 6-13: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of aldDH activity (U mg^{-1}) with respect to concentration of acetyl-CoA (mM) in the presence or absence of CoA-SH.

The intersection of the two lines below the X-axis indicates mixed inhibition with respect to CoA-SH where the value of K_i (dissociation constant for EI) is less than K_i' (dissociation constant for ESI) (Wharton and Eisenthal 1981). Thus, in the presence of CoA-SH, the K_m^{app} increases and the V_{max}^{app} decreases.

It was hypothesised that the tight binding of CoA-SH in the active site of the enzyme may be involved in the product inhibition. To confirm this, the kinetic parameters for CoA-SH for the AcAldDH in the reverse direction were determined, using the citrate synthase coupled assay described in Section 6.3.3. This assay was used due to the low concentration of CoA-SH present in the assay causing rapid curvature of the enzyme rate.

Activity (Substrate)	K_m (mM)	Standard Error	V_{max} (U mg^{-1})	Standard Error
AcAldDH reverse: (CoA-SH)	0.012	+/- 0.0002	11.60	+/-0.16

Table 6-5: Kinetic parameters of the reverse reaction of AcAldDH with respect to CoA-SH.

A relatively low K_m for CoA-SH was observed in the reverse direction. This K_m is lower than that for acetyl-CoA (0.024 mM). Tight binding of CoA-SH to the enzyme is a possible explanation for this difference and for the inhibition observed for this protein. As K_m is composed of both catalytic and binding constants, this hypothesis cannot be unequivocally confirmed.

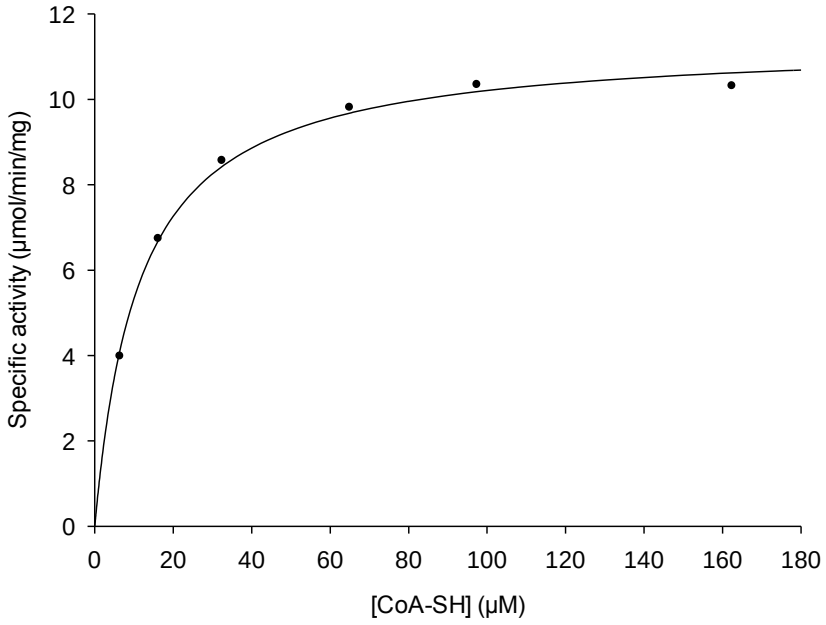


Figure 6-14: Michaelis-Menten plot of the reverse direction aldDH enzymic activity (U mg^{-1} of protein) against concentration of CoA-SH (µM) at a fixed concentration of NAD^+ (0.5 mM) & acetaldehyde (20 mM) for the purified AcAldDH enzyme.

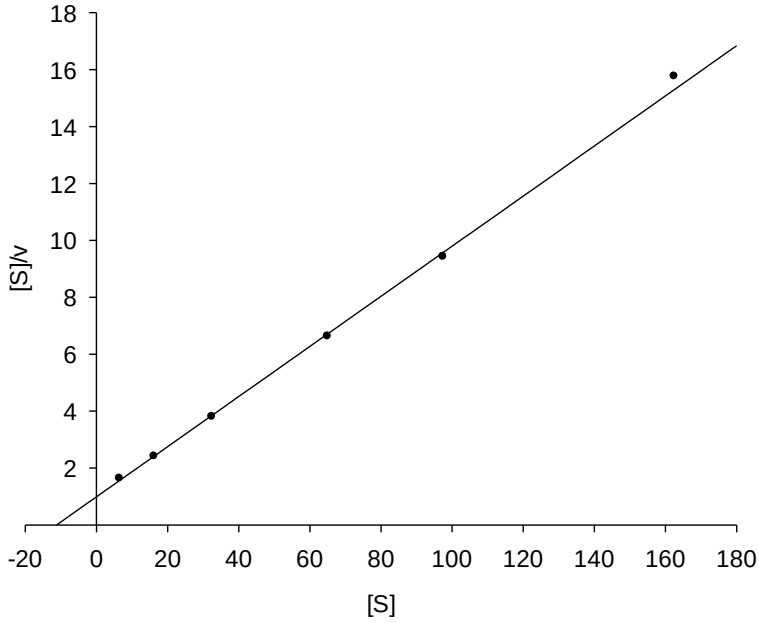


Figure 6-15: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of aldDH activity (U mg^{-1}) with respect to concentration of CoA-SH (µM).

6.4.6 Preliminary substrate specificity experiments

Preliminary experiments were carried out to assess the substrate specificity of the AcAldDH protein. AldDH activity was observed with both butyryl-CoA and propionyl-CoA. It was not possible to determine the concentration of these substrates, so accurate kinetic parameters could not be obtained. Assuming the concentration of these acyl-CoAs was comparable to that of the acetyl-CoA prepared at the same time,

the specific activities obtained with a fixed concentration of each acyl-CoA (0.28 mM) were compared.

Substrate	Specific activity (U mg^{-1})	% activity of Acetyl-CoA
Acetyl-CoA	52.0	100.0
Propionyl-CoA	14.3	27.5
Butyryl-CoA	4.7	9.0

Table 6-6: AldDH activity observed with different acyl-CoAs for the purified AcAldDH protein.

6.4.7 AcAldDH thermostability assays

Thermostability assays were carried out on the purified AcAldDH protein as described in Section 2.13.6. The protein was incubated in 100 µl aliquots at the desired temperatures.

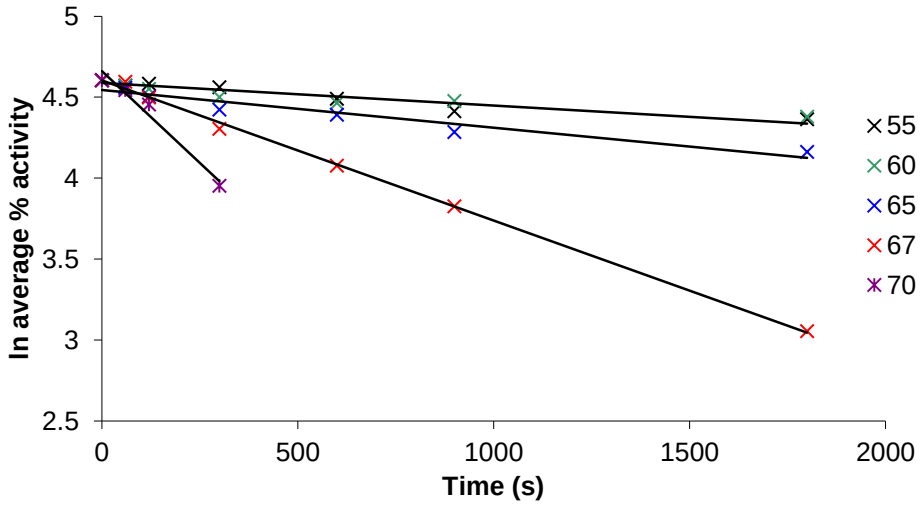


Figure 6-16: Plot of ln (% activity) versus time (s) at a defined temperature (°C) for aldDH activity of AcAldDH. Each data point displayed is an average of a minimum of 2 assays.

The half-life of the AcAldDH protein at 60°C was greater than 30 min. At 67°C the half-life was approximately 13 min.

6.4.8 AcAldDH optimum temperature assays

The dependence of enzyme activity on temperature was investigated with the purified AcAldDH protein as described in Section 2.13.5.

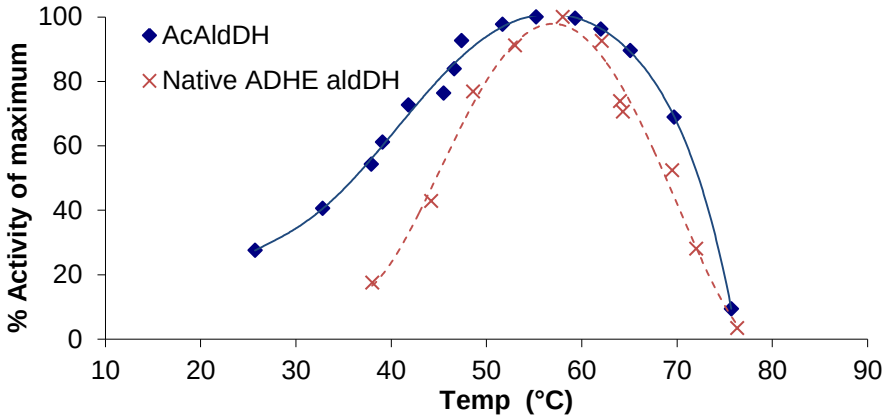


Figure 6-17: Plot of % activity of maximum against temperature (°C) for aldDH activity of AcAldDH and of native ADHE (Chapter 3).

The optimum temperature of the AcAldDH protein was approximately 57°C. This is almost identical to that observed for the aldDH activity of ADHE (58°C).

6.4.9 Oligomeric nature of AcAldDH

Gel filtration was carried out on purified AcAldDH (0.5 ml of 7.6 mg/ml solution loaded onto a pre-equilibrated GE Superdex 200 10/300 GL column run at a flow rate of 0.4 ml/min) in 50 mM EPPS pH 8.0, 5 mM reduced glutathione, 5 mM EDTA, and 10% (v/v) glycerol. SDS-PAGE was used to monitor the elution of the protein. The retention time of the major peak corresponding to the AcAldDH was compared to standard proteins to provide an estimate of the M_r.

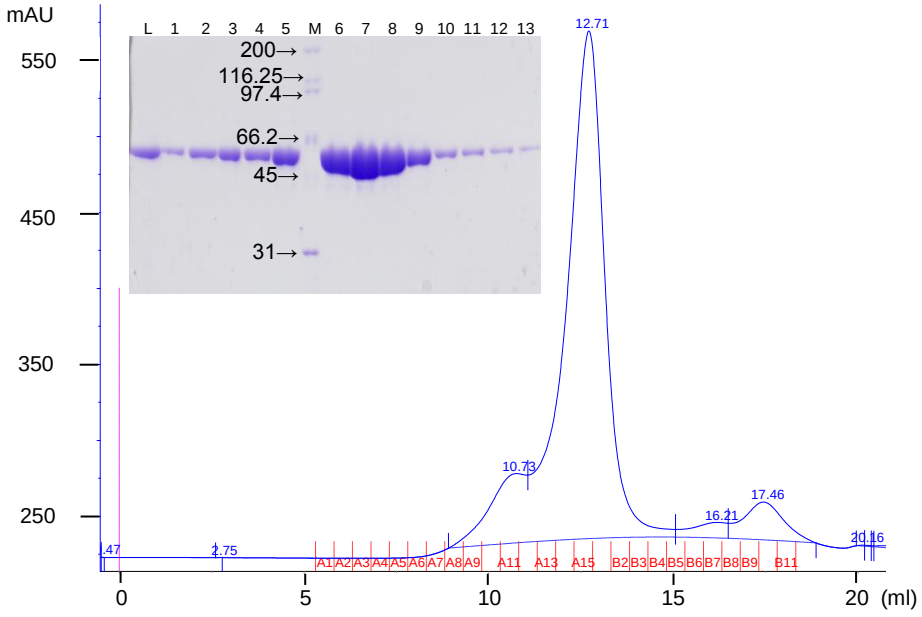


Figure 6-18: Summary chromatograph of an AcAldDH gel filtration run. Chromatograph line corresponds to A_{280nm} reading (mAU). SDS-PAGE analysis of peak fractions is shown (L = load, M = markers [M_r/1000], 1-13 = protein fractions from gel filtration A10-B7); the expected M_r of AcAldDH is 52,174.

Protein peak identity		$V_{elution}$ (ml)	$M_{r\ Observed}$	$M_{r\ Subunit}$	$M_{r\ Observed}$/ $M_{r\ Subunit}$
AcAldDH	peak 2	12.71	191,000	52,174	3.7

Table 6-7: Estimation of the M_r of AcAldDH by gel filtration.

The major protein peak corresponding to the AcAldDH protein eluted from the gel filtration column close to the predicted M_r of a tetramer.

6.5 Discussion

Two acetylating aldDH encoding genes were identified in the TM242 genome, *eutE* and a*cAldDH*. During characterisation of the native ADHE protein (Chapter 3), all fractions from an anion exchange run of a cell extract derived from a TM242 fermenter cell pellet were assayed for aldDH activity. No other peak of aldDH activity was observed other than that co-eluting with ADH activity (corresponding to ADHE). It is therefore assumed that these enzymes are not expressed during normal fermentative metabolism. AldDH assays performed during Chapter 9 on fermentation run samples also support this conclusion.

The EutE protein (25% identity, 43% similarity, 21% gaps compared to Fragment 1) appeared to be insoluble during expression experiments in both *E. coli* and *Geobacillus* strains. Although the addition of an MBP solubility tag did facilitate soluble protein expression, the protein produced appeared less catalytically active than the AcAldDH protein. Due to the issues surrounding its soluble expression and seemingly lower activity, it was decided to focus efforts on the AcAldDH protein.

The AcAldDH protein had a higher sequence identity than EutE to the aldDH domain of ADHE (43% identity, 59% similarity, 12% gaps to Fragment 1). This protein was readily expressed and purified from the *E. coli* expression strains used, by metal affinity chromatography. Running the protein on a gel filtration column showed that the protein appeared to associate to form tetramers. No oxygen sensitivity was observed for this protein in contrast to the aldDH domain of ADHE. Characterisation of the protein showed it to be active and stable around the growth temperature of *Geobacillus thermoglucosidasius.*

The kinetic parameters determined for the protein are shown in Table 6-8. The AcAldDH protein had a similar K_m for acetyl-CoA as that observed for the native ADHE protein. The K_m for NADH (in the presence of acetyl-CoA) was significantly lower for the AcAldDH than ADHE.

Protein	Activity	Substrate	Kinetic parameter	Parameter value	Units
AcAldDH (purified)	AldDH	Ac-CoA	K_m	0.024	mM
			$V_{max\ (adjusted)}$	37.5	U mg^{-1}
	AldDH	NADH	K_m	0.042	mM
			$V_{max\ (adjusted)}$	34.5	U mg^{-1}
ADHE (TM242 cell extract)	AldDH	Ac-CoA*	K_m	0.019	mM
			$V_{max\ (adjusted)}$	1.24	U mg^{-1}
	AldDH	NADH	K_m	0.16	mM
			$V_{max\ (adjusted)}$	1.0	U mg^{-1}

Table 6-8: Comparison of kinetic parameters for purified AcAldDH (produced in BL21) and the aldDH of ADHE in TM242 cell extracts. *Kinetic parameters for the aldDH domain with respect to acetyl-CoA were determined by fitting the data to the substrate inhibition equation as described in Section 2.13.7.3. V_{max} values reported are adjusted for saturation with the fixed substrate.

The catalytic parameters determined for the AcAldDH protein are shown below:

Activity	Substrate	V_{max} (U mg^{-1})	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}\bullet s^{-1}$)
AcAldDH					
aldDH	Acetyl-CoA	37.5	0.024	32.6	1358
	NADH	34.5	0.042	30.0	714

Table 6-9: Summary of the kinetic parameters determined for the purified AcAldDH protein.

The V_{max} values obtained for the native protein in cell extracts and for the purified AcAldDH protein cannot be directly compared. The enzyme appears to be as, if not more, catalytically active than the aldDH activity of ADHE with a similar K_m for acetyl-CoA. The protein was shown to be catalytically active with butyryl and propionyl–CoAs, but a significantly higher activity was observed with acetyl-CoA. The native ADHE appeared to be less specific for acetyl-CoA (Chapter 3).

There was a discrepancy in V_{max} values determined between Sections 6.4.4 and 6.4.5 (37.5 and 57.7 U.mg^{-1}, respectively). Differences in protein purity between the different enzyme preparations used are a likely explanation for these observed differences.

The AcAldDH protein appeared to be inhibited by the products of the catalysed reaction. Inhibition was shown to be mixed with respect to CoA-SH, which had the strongest inhibitory effect. Under mixed inhibition, the inhibitor can bind to the free enzyme (E) and the substrate-bound enzyme complex (ES). When the data were plotted on a Hanes-Woolf plot, the lines intersect below the X-axis. This shows the

inhibitor binds the free enzyme more strongly than the ES complex as K_i is lower than K_i'. This is illustrated in Figure 6-19.

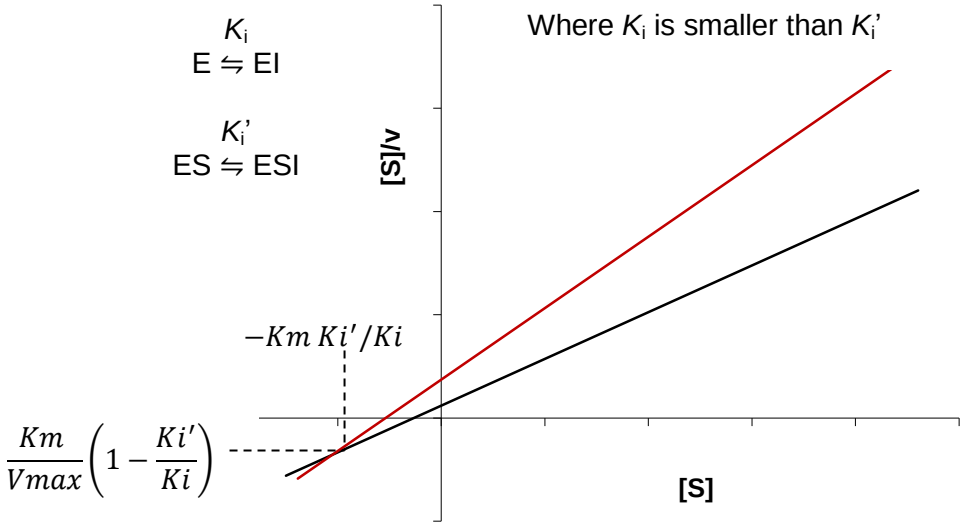


Figure 6-19: Representative Hanes-Woolf plot ([S]/v vs. [S]) for the variation of enzyme activity with respect to concentration of S, in the presence (red line) or absence (black line) of a fixed concentration of a mixed inhibitor. Figure based on Figure 6.10 from Wharton and Eisenthal (1981).

Both acetaldehyde and CoA-SH were shown to inhibit the enzyme. The K_m for CoA-SH in the reverse direction was determined to be 0.012 mM; this is lower than that determined for acetyl-CoA in the forward direction (0.024 mM). The inhibition observed could be caused by inefficient release of these products from the active site of the enzyme.

During preliminary experiments, the effect of assaying the aldDH activity of AcAldDH in the presence of the Fragment 11 protein was tested. Neither the rate of reaction nor the curvature of the assay was affected by the presence of excess Fragment 11. This suggested the acetaldehyde produced was either not efficiently released by the AcAldDH, or not effectively bound by the ADH protein. It was hypothesised that the positioning the ADH domain close to the AcAldDH through a protein fusion, may help over-come the product inhibition which had been observed for this enzyme. A protein fusion between the AcAldDH and Fragment 11 was also of interest due to the potential for improving ADHE protein activity in terms of ethanol production.

The aldDH domain of ADHE was unable to be resolved structurally during this project. The reasonably high sequence identity with the aldDH domain of ADHE implies a high resolution AcAldDH structure may provide useful information regarding the ADHE

protein. A high resolution crystal structure was also considered desirable as it could prove informative as to potential targets for modification of this protein to alleviate product inhibition. The protein could be readily purified to the high yields required for structural characterisation, which is described in Chapter 7.

In summary, genes encoding two active acetylating aldDH proteins have been successfully identified within the TM242 genome. The catalytically active and readily expressed AcAldDH protein has been characterised. The enzyme was stable and optimally active at the required temperatures for use in the TM242 organism. Product inhibition was observed for the protein with respect to the reaction products. Subsequent chapters describe the structural resolution of the AcAldDH protein and the production of fusion proteins consisting of Fragment 11 and the AcAldDH.

7 CRYSTALLISATION OF THE ACETYLATING ALDEHYDE DEHYDROGENASE PROTEIN (ACALDDH)

7.1 Introduction

This chapter details the determination of two crystal structures for the AcAldDH protein identified and characterised in Chapter 6. One structure at high resolution is reported in the absence of substrates or products, with a lower resolution structure in the presence of NAD^+ and acetyl-CoA also being obtained.

The structure of this protein is of interest for a number of reasons. Firstly, it will contribute to an understanding of the AcAldDH mechanism, particularly the product inhibition observed; it was hoped that key residues potentially interfering with product release may be identified, allowing optimisation of catalytic activity. Optimisation of catalysis was also desirable as the AcAldDH was used in the creation of an artificial ADHE protein through fusion with the isolated ADH domain known as Fragment 11 (described during Chapter 8). Product inhibition remained following the fusion of the two domains, and so manipulation of the AcAldDH may also be used to improve the productivity of fusion proteins described subsequently.

Secondly, ordered substrate binding in a "bi-uni-uni-uni ping-pong" scheme has been suggested for other acetylating aldDH enzyme mechanisms (Shone and Fromm 1981; Smith and Kaplan 1980); this suggests that a shared binding site may exist for NADH and acetyl-CoA, which may be the origin of the product inhibition observed during Chapter 6 for this enzyme. Structural investigation of substrate binding was carried out to test this hypothesis.

As the aldDH domain of the ADHE protein shares significant sequence identity with the AcAldDH protein (43% identity, 59% similarity, 12% gaps), further development of the structural understanding of similar proteins was appealing. Isolated acetylating aldDH proteins are not extensively described in the literature, so additional structural information may also prove valuable for future investigations.

7.2 AcAldDH methods

Unless otherwise stated, crystallographic methods used here were the same as those described in Chapter 5.

1 M MMT buffer at pH 4 and pH 9 (adjusted with HCl or NaOH) was prepared according to the PACT screen manufacturer's instructions and mixed to obtain the desired pH buffer for crystallisation. The buffer comprised DL-malic acid, MES and Tris base mixed 1:2:2 respectively.

7.2.1 Soaking experiments

Attempts to obtain crystal structures of the AcAldDH protein with substrates/products bound in the active site were made. Crystals were transferred using a cryo-loop from the drop in which they had grown, into a fresh drop containing the well solution with the appropriate concentration of cryo-protectant and substrate/product of interest. Crystals were then mounted onto the collection apparatus within a stream of nitrogen gas to freeze the crystals.

7.2.2 Room-temperature diffraction

Where crystals being screened did not diffract well under freezing conditions, room-temperature diffraction was attempted. The Micro RT™ system (MiTeGen, USA) was used to mount the crystals as shown in Figure 7-1. This prevents the crystals drying out due to the reservoir of well solution at the top of the tube allowing vapour diffusion to keep the crystals hydrated.

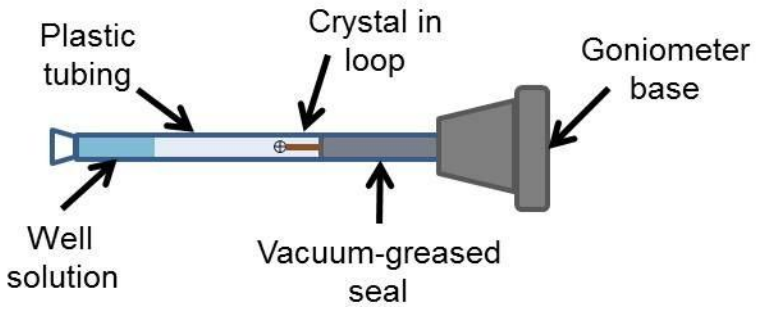


Figure 7-1: The Micro RT™ system for room-temperature diffraction (MiTeGen 2012).

7.3 AcAldDH crystallisation results

Crystallisation of the AcAldDH protein was attempted in the absence and in the presence of substrates. Results obtained for the two conditions will be reported separately in the following sections.

7.4 AcAldDH protein crystallisation without substrates

7.4.1 Crystallisation conditions

The purified His-tagged AcAldDH protein (described in Chapter 6) dialysed into 50 mM Tris-HCl pH 8.0, 150 mM NaCl, was screened with structure screen I & II, PGA, JCSG-plus and PACT screens (Molecular Dimensions, UK), at a protein concentration of 8.9 mg/ml. The ratio of well solution to protein was varied between the three wells of the screens (300 nl drop size); mixtures of 1:1, 2:1 and 1:2 protein: well solution (v/v) were used.

The best hit obtained was for the PACT screen, well A7 (0.1 M sodium acetate pH 5.0, 0.2 M NaCl, 20% (v/v) PEG 6000) for the 2 protein: 1 well solution (v/v) sample.

To obtain the most suitable crystals for X-ray diffraction, optimisation of the conditions employed a variety of concentrations of different metal salts (NaCl, $MgCl_2$, LiCl & $CaCl_2$) and various gradients of PEG 6000 and pH.

Crystals producing the best diffraction data were grown at a protein concentration of 9.25 mg/ml, diluted 2:1 (v/v) with well solution (0.1 M sodium acetate pH 4.8, 0.2 M $MgCl_2$, 15% (v/v) PEG 6000). Crystals took 2-4 weeks to appear at 16°C, in which time a "skin" had formed on the protein drops; this had to be peeled back using a needle to allow crystals to be removed. Representative crystal forms are shown in Figure 7-2. Glycerol (30%, v/v) was required as a cryo-protectant when the crystals were frozen for data collection.

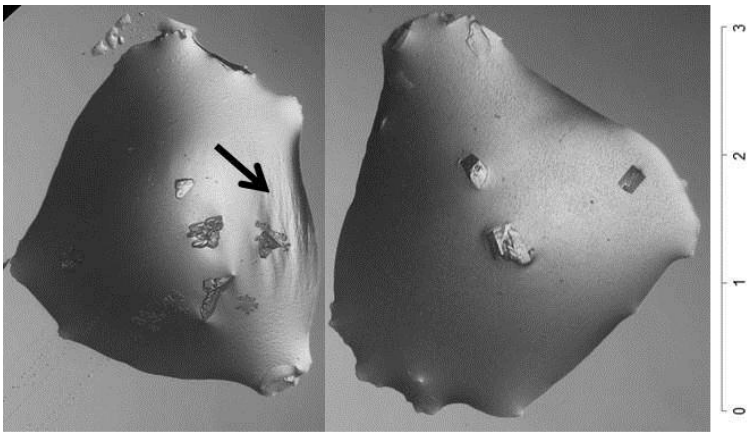


Figure 7-2: AcAldDH (no substrate) representative crystal forms (scale shown = mm), indications of skin on the drop can be observed (indicated with an arrow) in the left hand image.

7.4.2 Data collection and molecular replacement

Data were collected at the Diamond light source beam-line IO4 at a wavelength of 0.97949 Å. Data were processed using the HKL2000/0.98 software package. Statistics of data collection are shown in Table 7-1.

Parameter	Value
Unit cell dimensions	a = 98.579 Å, b = 109.120 Å, c = 196.243 Å, α=β=γ = 90.00°
Space group	$P2_12_12_1$
R_{merge}	0.136 (0.684)
Completeness	99.9 (98.9)%
I/σ(I)	15.4 (2.23)
Multiplicity	7.1 (5.5)
Number of reflections observed	2590618
Number of unique reflections	123787 (6036)
Resolution	50.0-2.1 Å (2.14-2.10)

Table 7-1: AcAldDH data collection statistics (numbers in brackets for highest resolution bin).

A range of crystal forms in various space groups had been previously tested in-house prior to this data collection. One of the preliminary $p2_12_12_1$ structures was used as a starting model for the Diamond dataset.

The preliminary structure used was based on the BALBES (Long et al. 2008) suggested structure of *Listeria monocytogenes* aldehyde dehydrogenase (PDB code 3K9D) for molecular replacement (BALBES solution was 99% likely). The amino acid sequence of 3K9D is 53% identical and 73% similar (1% gaps) to the non-His-tagged AcAldDH sequence. The preliminary structure was rigid-body refined into the Diamond data and the refinement process started as previously described.

7.4.3 Data refinement

Following the rigid-body refinement in Refmac5, the R factor was 0.27916 and R_{free} was 0.27407.

The model was subjected to several rounds of refinement and validation; 4 AcAldDH proteins were visible in the unit cell (molecules A-D). Several different non-protein molecules present in the crystallisation condition, as well as 1251 waters, were added to regions of un-modelled density (full list in Table 7-2), with the model refinement process continuing until no further improvement in fit to the density was observed. Dual conformations were observed for 39 residues in the final structure and were

therefore modelled with partial occupancy. Statistics for the refined final model are shown in Table 7-3.

Molecule	Number observed
Water	1251
Glycerol	30
PEG fragments	3
Acetate	12
Chloride	4
Magnesium	4

Table 7-2: A summary of any additional molecules visible in the AcAldDH crystal structure.

Reasonable electron density was observed for residues 25 to 476 of all AcAldDH molecules; however, the N-terminal 25 amino acids and C-terminal 12 amino acids proved difficult to resolve due to unresolvable or poor electron density. Ten difference map peaks ($F_o - F_c$) above σ level 5.0 remained at the end of refinement, 9 of which were associated with the N-termini.

Parameter	Value
Number of reflections used	117457 (5% test set)
Number of protein atoms	13967
Number of non-solvent molecules	53
Number of solvent molecules	1251
Mean B-factors (Å^2)	
Overall	25.9
Protein main chain	22.7
Side chains	27.1
Water molecules	35.5
R	0.1551
R_{free}	0.2085
RMS Bond Length	0.0193 Å
RMS Bond Angle	1.9408°

Table 7-3: Final refinement statistics for the AcAldDH crystal structure.

Evaluation of the structure was carried out using MolProbity (Chen et al. 2010).

Parameter	Value
Ramachandran favoured	98.4% (1866/1896)
Ramachandran allowed	100% (1896/1896)
Ramachandran outliers	0
MolProbity score	1.80 (90th percentile)
Clashscore	10.96 (82nd percentile)

Table 7-4: Validation results for the AcAldDH structure from MolProbity.

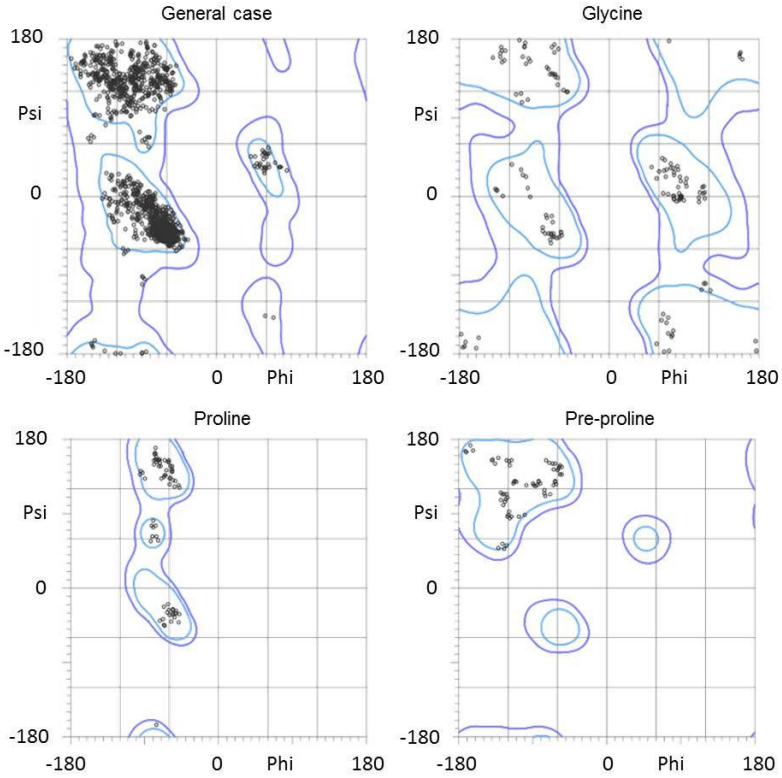


Figure 7-3: MolProbity Ramachandran analysis of the AcAldDH structure.

Positive electron density was observed at the terminal S_γ of Cys_{273} in all AcAldDH monomers; the residue was replaced with an S-cysteinesulfinic acid during refinement, and this significantly improved the density fit as shown in Figure 7-4. High-energy synchrotron radiation may allow oxidation of this residue through radiation damage. This modified residue has been observed in over 180 other high-resolution protein crystal structures within the PDB (RCSB 2012).

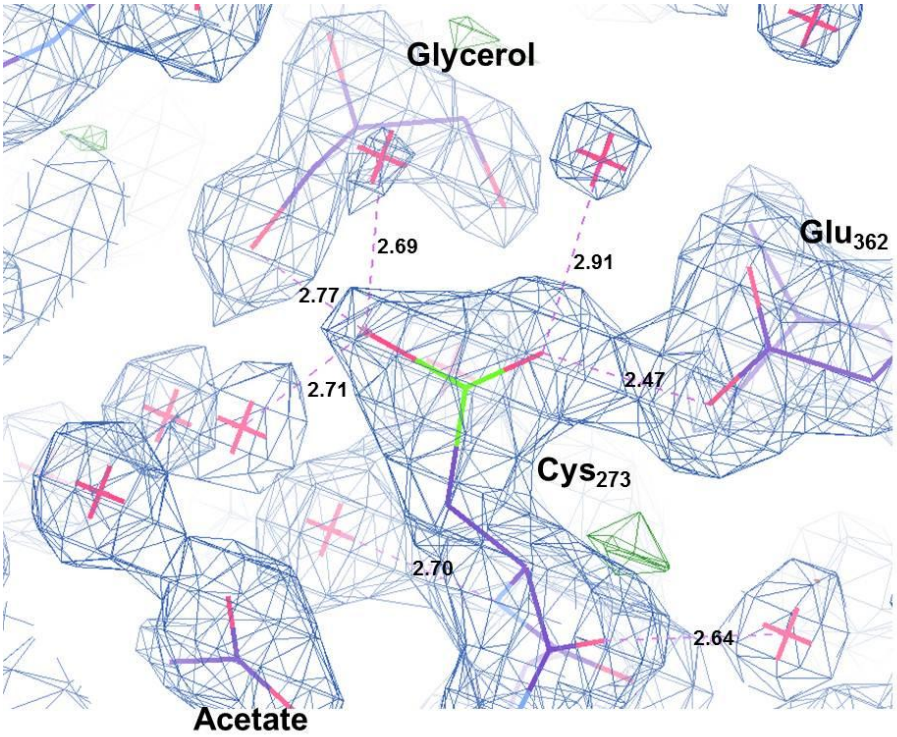


Figure 7-4: Electron density map for the modified Cys_{273} (S-cysteinesulfinic acid), demonstrating the improved fit and lack of difference density (residue from AcAldDH molecule A shown). Blue map = $2F_o$ - F_c (where F_o = $F_{observed}$, F_c = $F_{calculated}$, and F = structural factor), σ level = 1.10, negative (red) and positive (green) difference map (F_o – F_c), σ level = 3.0. Molecules are shown in stick form, with carbons in purple. Red crosses represent water molecules.

Sequence alignments between AcAldDH and the aldDH domain of ADHE showed that this cysteine residue aligns with the conserved catalytic cysteine (Cys_{257} in ADHE) described in the Introduction to this thesis; it is also located proximal to the conserved Glu_{362} residue (part of the EKLSP motif). This Glu has been suggested to be involved in the initiation of catalysis through the removal of the Sγ proton from Cys_{273} (Chen et al. 2004).

7.4.4 Interpretation of the AcAldDH structure

Two structural domains were observed for the AcAldDH protein. The CATH structural comparison tool was used to determine the structural classification of these domains. The N-terminal domain was shown to be part of the 3.40.605.10 superfamily (aldehyde dehydrogenase; chain A, domain 1) and the C-terminal domain was shown to be part of the 3.40.309.10 superfamily (aldehyde dehydrogenase; chain A, domain 2). The two domains both have 3-layer (αβα) sandwich topologies. The structure of the modelled aldDH of ADHE discussed in Chapter 5 also appeared to be made up of the same structural domains.

7.4.4.1 Overview

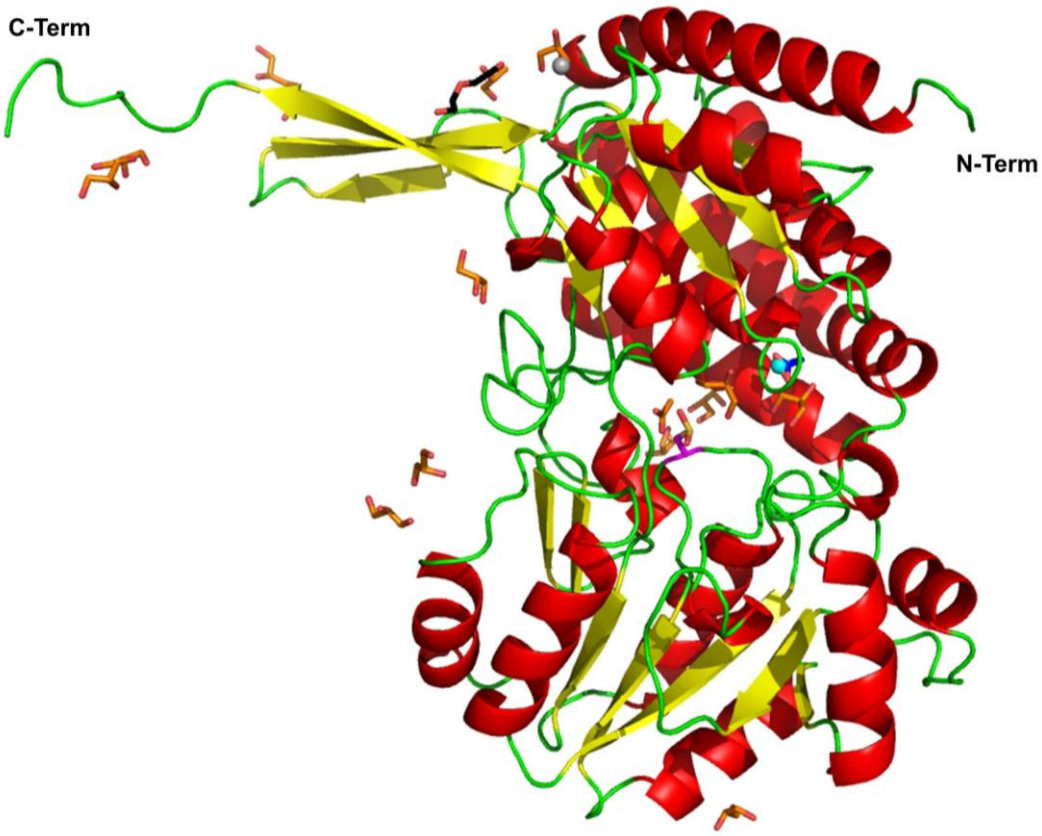


Figure 7-5: Cartoon overview of AcAldDH chain A. Red spirals = α-helices and yellow arrows = β-strands. Ligands are shown in stick representations, the glycerols and acetates with orange carbon atoms and a PEG molecule with black carbons. A magnesium ion is shown in grey and the acetate/chloride molecule is coloured blue. The catalytic cysteine (Cys_{273}) is shown in purple.

7.4.4.2 Structural alignment between AcAldDH and homologous structures

Structural alignments between AcAldDH and structures suggested by BALBES as good molecular replacement models were carried out using the DaliLite pairwise protein comparison tool (EMBL) (Holm and Park 2000). The structures suggested by BALBES were the same structures used to create the aldDH domain model during Chapter 5.

PDB code (chain)	Source organism	Protein description	Aligned residues	Seq ID	Z-score	RMSD (Å)
3K9D (A)	*Listeria monocytogenes*	Probable aldehyde dehydrogenase	450	55%	60.8	1.1
3MY7 (A)	*Vibrio parahaemolyticus*	aldDH domain of an ADHE	404	45%	49.4	2.0

Table 7-5: Results of structural alignment analysis using DaliLite. Z-Score describes the quality of the alignment (if >20 structures are considered homologous). RMSD is used as a measure of deviation in terms of distance of the Cαs in the aligned structures.

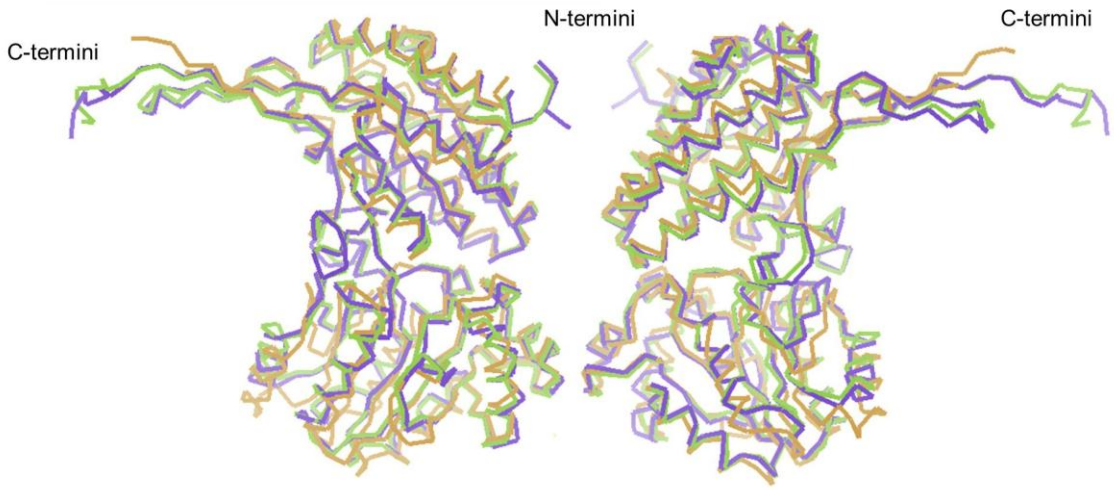


Figure 7-6: Cα traces of aligned A chains of AcAldDH (purple), 3K9D (green) and 3MY7 (orange); "front and back" views are shown. Visible termini are indicated.

The AcAldDH structure aligns well with 3K9D; a relatively low RMSD value was observed between the two structures and few differences were observed in the structural superimposition (Figure 7-6). The alignment with 3MY7 was not as good, but it should be noted that several loop regions are missing in the 3MY7 structure which may affect the quality of the alignment. The protein sequence of AcAldDH was also more similar to 3K9D than to the 3MY7 sequence.

7.4.5 Other molecules in the structure

As reported in Table 7-2, several additional molecules were visible within the AcAldDH structure. Various crystallisation condition components were fitted into the observed density and those appearing to fit best were refined into the final model. An octahedrally-coordinated ion, interacting with Asp_{215} and four water molecules, was observed at the surface of all the protein chains, and a Mg^{2+} ion was refined into this density and appeared to fit reasonably (Figure 7-7 i). Strong difference density was observed beside each chain near to $Lys_{92,}$ Lys_{95} and Ser_{145}; a variety of possible molecules were modelled into this region and the model refined to evaluate the fit. The best refinement statistics were obtained when a combination of a chloride ion and an acetate were refined into this region. This suggests that a variety of molecules may be present here interacting with the surrounding residues (Figure 7-7 ii). These additional molecules/ions appear to be a consequence of crystallisation conditions, rather than playing any key structural or catalytic role within the protein.

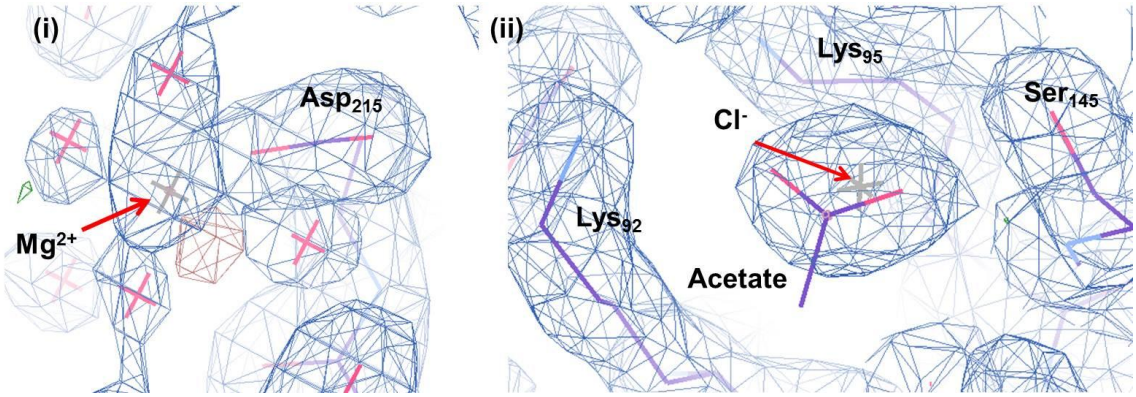


Figure 7-7: Electron density map for AcAldDH molecule D focussed on: (i) the Mg^{2+} and (ii) the acetate/Cl^- which were refined into the final model. Blue map = $2F_o$ - F_c (where F_o = $F_{observed}$, F_c = $F_{calculated}$, and F = structural factor), σ level = 1.10; negative (red) and positive (green) difference map (F_o – F_c), σ level = 3.0. Molecules are shown in stick form, with carbons in purple. Red crosses represent water molecules.

Twenty four glycerol molecules as well as three PEG fragments were also observed in the structure around the surface of the protein. Glycerol was used as a cryo-protectant and PEG 6000 was present in the well solution, and therefore the presence of these molecules surrounding the protein was expected.

The predicted substrate-binding cleft of the protein chains A-C appeared to contain two additional glycerols and one acetate molecule close to Cys_{273} (Figure 7-4 and Figure 7-8). These are also located in the vicinity of His_{166} which is shown in Section 7.5.5 to be close to the region with which the ADP moiety of NADH and acetyl-CoA may interact. Only density for the acetate was observed in the active site of chain D with waters appearing to replace the glycerols; this difference may be due to limited electron density observed for this region. The presence of these molecules supports the concept of a polar binding-cleft into which the non-homologous regions of NADH and acetyl-CoA may bind.

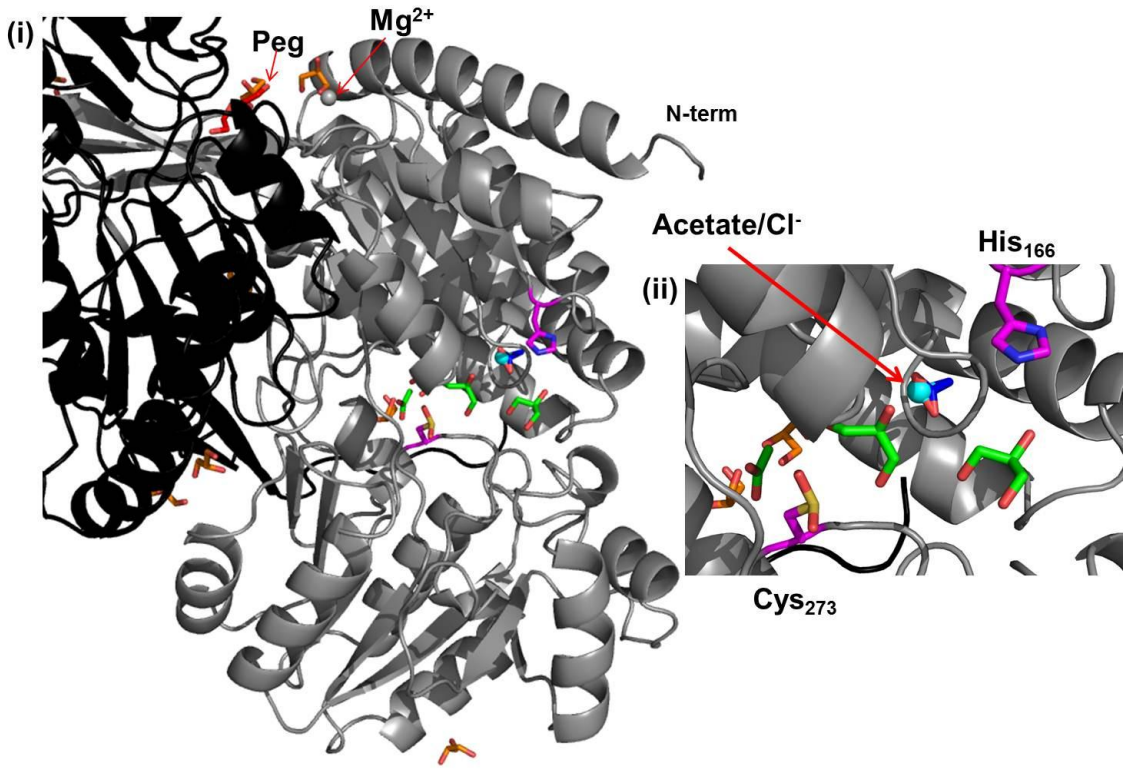


Figure 7-8: (i) AcAldDH cartoon structure chain A (grey) with other molecules in the structure visible (chain B is shown in black). Chain A Cys_{273} and His_{166} are shown with carbon atoms coloured purple. Ligands are represented in stick form with non-active site glycerol and acetate molecules carbon atoms shown in orange. A PEG molecule is shown with red carbon atoms and the acetate-Cl^- is shown in blue. The two glycerol molecules and the acetate present in the active site are shown with green carbon atoms. (ii) A close-up view of the active-site region of chain A.

7.4.6 Multimeric assembly

Electron density corresponding to the N-terminal region of the protein (aa 1-24) could not be resolved for the AcAldDH chains; the N-termini of chains A and C, and B and D, appear to be closely interacting. Despite attempts during model building and refinement, the orientation of these termini could not be resolved; the chains were therefore truncated such that one residue of the AcAldDH protein (Met_{24}) was missing from the structure. The remaining residues correspond to the histidine tag incorporated to facilitate purification. Had time permitted, it would have been preferable to resolve the AcAldDH protein structure following cleavage of the tag to determine the native positioning of these terminal residues. Up to 12 amino acids were unresolved at the C-terminus of the protein (VEGAQKEPALTK-); as close interactions between the C-terminal regions were observed in the structure, it is likely that these additional amino acids are not required for dimer formation, and may be highly mobile within the structure and thus unresolvable from the collected data.

Gel filtration analysis carried out during Chapter 6 suggested that the AcAldDH protein may form tetrameric assemblies. A tetrameric assembly was also observed during crystallographic analysis as shown in Figure 7-9. Analysis of the structure using PISA showed the dimerization interaction between monomers A and B, or C and D, as identified in Figure 7-9, to be the most physiologically relevant of the interactions (data summarised in Table 7-6). This interaction appears to form as a result of a "wrapping round" of the protein C-termini to form stable dimeric units as shown in Figure 7-10. The formation of a "dimer of dimers" (interaction between dimer A-B and C-D in Figure 7-9) would appear to be a relatively weak interaction compared to the formation of these dimeric units. A summary of the residues forming interactions between proteins A and B is shown in Table 7-7 and Table 7-8. For all chains, the predicted catalytic "cleft" containing the Cys_{273} residue is positioned facing inwards towards the N-terminal region of the tetramer.

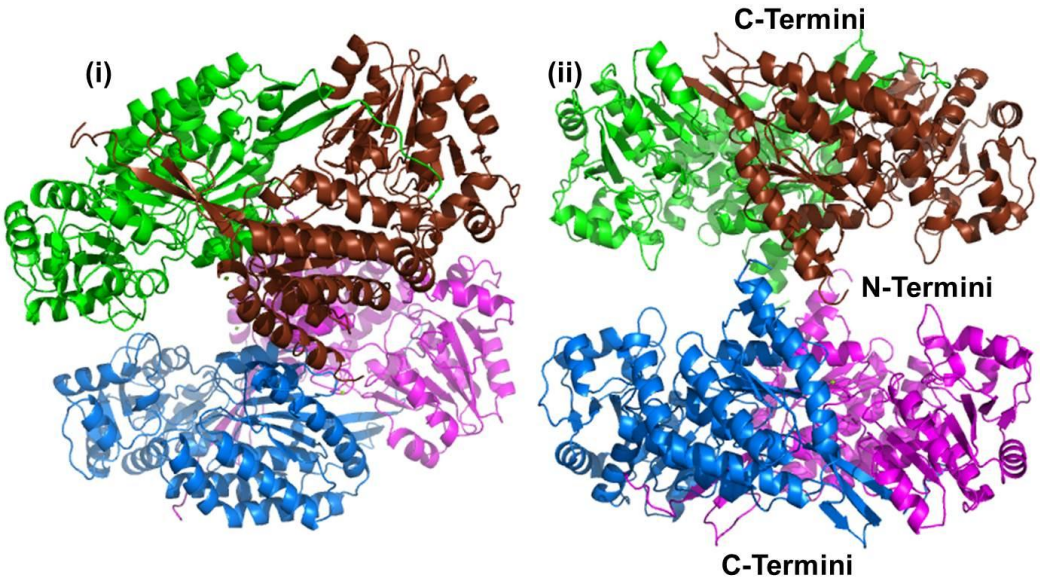


Figure 7-9: Cartoon diagram showing the interactions between AcAldDH proteins within the crystal. (i) = "upper" and (ii) = "side-on" views. Green = molecule A, brown = molecule B, purple = molecule C, and blue = molecule D. Terminal regions are indicated in (ii).

Interaction #	# interfacing residues		Interface area (Å^2)	Δ^iG (kcal/mol)	Δ^iG P-value	#HB	#SB	CSS
	A	B						
1	86	85	2983.3	-24.4	0.192	56	14	0.430
	C	D						
2	82	81	2912.9	-22.2	0.212	57	8	0.430
	A	C						
3	9	11	455.7	3.4	0.896	15	13	0.000
	B	D						
4	11	11	462.9	3.9	0.903	11	12	0.000

Table 7-6: Summary of the PISA analysis results for AcAldDH multimer formation. Δ^iG = solvation free energy gain upon interface formation (-ve = hydrophobic interface). P-value = probability that the decrease in ^{i}G through the interface is not random, where P>0.5 = non-specific and P<0.5 shows interface that may be considered interaction specific. HB = hydrogen bonds formed, SB = salt bridges formed. CSS = complexation significance score, which is a measure of interface significance in terms of complex formation (the higher the value between 0 and 1, the more significant the interface).

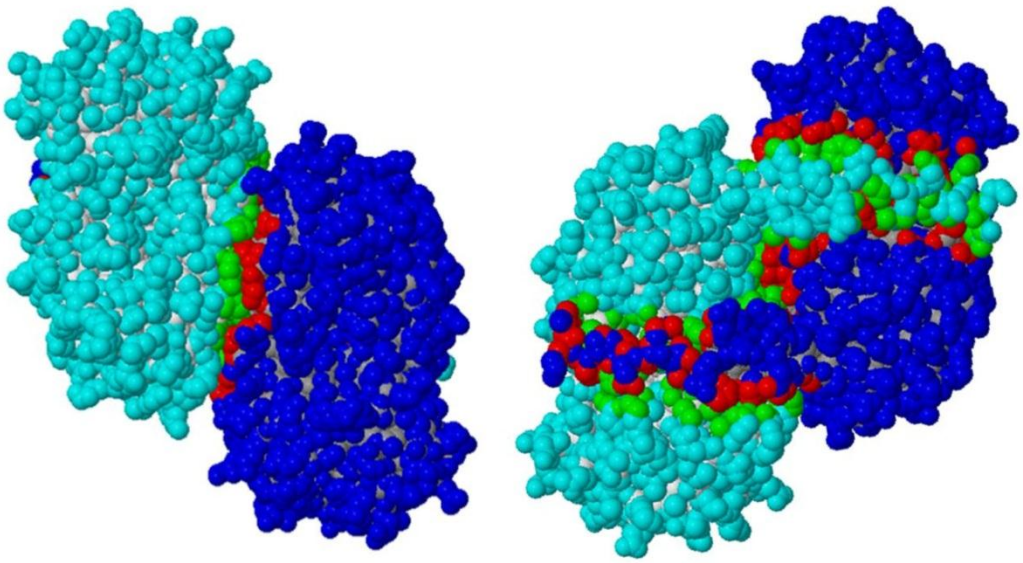


Figure 7-10: PISA interface space-filling diagram for interaction 1 (front and back views shown). Dark blue = molecule A, light blue = molecule B. Interface residues are shown in red for A and green for B.

Salt Bridge Forming Residues							
#	Molecule A	D (Å)	Molecule B	#	Molecule A	D (Å)	Molecule B
1	GLU407 [OE2]	3.93	ARG118 [NH2]	8	ARG118 [NH1]	3.18	GLU407 [OE1]
2	GLU472 [OE1]	3.43	LYS256 [NZ]	9	ARG118 [NH2]	3.35	GLU407 [OE1]
3	GLU472 [OE2]	3.17	LYS256 [NZ]	10	ARG118 [NH2]	3.73	GLU407 [OE2]
4	ASP453 [OD1]	2.89	ARG465 [NH1]	11	ARG465 [NH1]	2.79	ASP453 [OD1]
5	ASP453 [OD1]	3.33	ARG465 [NH2]	12	ARG465 [NH2]	3.3	ASP453 [OD1]
6	ASP91 [OD1]	2.92	LYS475 [NZ]	13	LYS256 [NZ]	3.72	GLU472 [OE1]
7	ASP91 [OD2]	3.48	LYS475 [NZ]	14	LYS256 [NZ]	3.24	GLU472 [OE2]

Table 7-7: Summary of the residues in the AcAldDH interaction 1 forming salt bridges according to PISA analysis ([X] = atom identity from annotated residue, D = distance between interacting atoms).

Hydrogen Bond Forming Residues							
#	Molecule A	D (Å)	Molecule B	#	Molecule A	D (Å)	Molecule B
1	TYR230 [O]	3.47	SER232 [N]	**29**	ASN434 [ND2]	2.89	ASN122 [O]
2	TYR230 [O]	2.8	GLY233 [N]	**30**	GLY233 [N]	2.74	TYR230 [O]
3	TYR230 [O]	3.85	LYS234 [N]	**31**	LYS234 [N]	3.83	TYR230 [O]
4	ILE450 [O]	3.86	TYR237 [OH]	**32**	SER232 [N]	3.59	TYR230 [O]
5	GLU472 [OE1]	3.43	LYS256 [NZ]	**33**	LYS227 [NZ]	3.27	GLY233 [O]
6	GLU472 [OE2]	3.17	LYS256 [NZ]	**34**	ASN448 [ND2]	3.02	GLY233 [O]
7	MET473 [O]	3.48	GLN263 [NE2]	**35**	ARG118 [NH1]	3.18	GLU407 [OE1]
8	ARG465 [O]	2.99	ILE418 [N]	**36**	ARG118 [NH2]	3.35	GLU407 [OE1]
9	ALA467 [O]	2.93	VAL420 [N]	**37**	LYS464 [NZ]	2.97	ILE411 [O]
10	ASN122 [O]	2.91	ASN434 [ND2]	**38**	LYS464 [NZ]	2.65	LYS413 [O]
11	ASN462 [OD1]	3.1	ALA445 [N]	**39**	ARG465 [N]	2.94	GLY416 [O]
12	GLY233 [O]	2.87	ASN448 [ND2]	**40**	ALA467 [N]	2.88	ILE418 [O]
13	ILE463 [O]	2.74	SER452 [OG]	**41**	GLY469 [N]	2.86	VAL420 [O]
14	SER452 [OG]	3.29	ILE463 [N]	**42**	ARG471 [NH2]	3.46	GLY430 [O]
15	LYS413 [O]	2.74	LYS464 [NZ]	**43**	ARG471 [NE]	2.95	ALA431 [O]
16	ILE411 [O]	3.34	LYS464 [NZ]	**44**	ARG471 [NH2]	3.37	ALA431 [O]
17	GLY416 [O]	2.87	ARG465 [N]	**45**	VAL470 [N]	2.96	THR432 [O]
18	ASP453 [OD1]	2.89	ARG465 [NH1]	**46**	ARG471 [N]	3.39	THR432 [O]
19	ILE418 [O]	2.87	ALA467 [N]	**47**	ARG471 [NH2]	3.39	VAL433 [O]
20	VAL420 [O]	2.89	GLY469 [N]	**48**	ARG471 [NH2]	3.17	ASN434 [OD1]
21	THR432 [O]	2.87	VAL470 [N]	**49**	ARG471 [NH1]	2.8	ASN434 [OD1]
22	THR432 [O]	3.38	ARG471 [N]	**50**	ILE463 [N]	3.4	SER452 [OG]
23	ALA431 [O]	3.00	ARG471 [NE]	**51**	ARG465 [NH1]	2.79	ASP453 [OD1]
24	ASN434 [OD1]	2.78	ARG471 [NH1]	**52**	ALA445 [N]	2.97	ASN462 [OD1]
25	GLY430 [O]	3.36	ARG471 [NH2]	**53**	SER452 [OG]	2.72	ILE463 [O]
26	VAL433 [O]	3.37	ARG471 [NH2]	**54**	ILE418 [N]	2.65	ARG465 [O]
27	ALA431 [O]	3.2	ARG471 [NH2]	**55**	VAL420 [N]	2.93	ALA467 [O]
28	ASP91 [OD1]	2.92	LYS475 [NZ]	**56**	LYS256 [NZ]	3.24	GLU472 [OE2]

Table 7-8: Summary of residues forming hydrogen bonds in the AcAldDH interaction 1 according to PISA analysis ([X] = atom identity from annotated residue, D = distance between interacting atoms).

It should be noted that interactions between AcAldDH molecules may be further stabilised by the missing regions at the N and C termini that were not resolved in this structure.

7.4.7 Modelled NAD^+ binding

No structurally similar proteins to AcAldDH were identified within the PDB with either acetyl-CoA or CoA-SH bound; a structurally similar protein with a nicotinamide cofactor bound was identified (3JZ4) (Langendorf et al. 2010). This semi-aldehyde dehydrogenase was superimposed onto the AcAldDH structure (shown in Figure 7-11) and the structural alignment was evaluated using DaliLite as shown in Table 7-9. The $NADP^+$ from this structure was then overlaid onto AcAldDH to provide an indication of the NAD^+ binding site. The substrate-binding regions of the proteins aligned well, but significant variation was observed for some of the loops on the N-terminal side.

PDB code (chain ID)	Source organism	Protein description	Aligned residues	Seq ID	Z-score	RMSD (Å)
3JZ4 (A)	*E. coli*	Succinic semi-aldehyde dehydrogenase	385	19%	31.0	2.8

Table 7-9: Results of the structural alignment analysis between AcAldDH and 3JZ4 using DaliLite. Z-Score describes the quality of the alignment (if >20 structures are considered homologous). RMSD is used as a measure of deviation in terms of distance of the Cαs in the aligned structures.

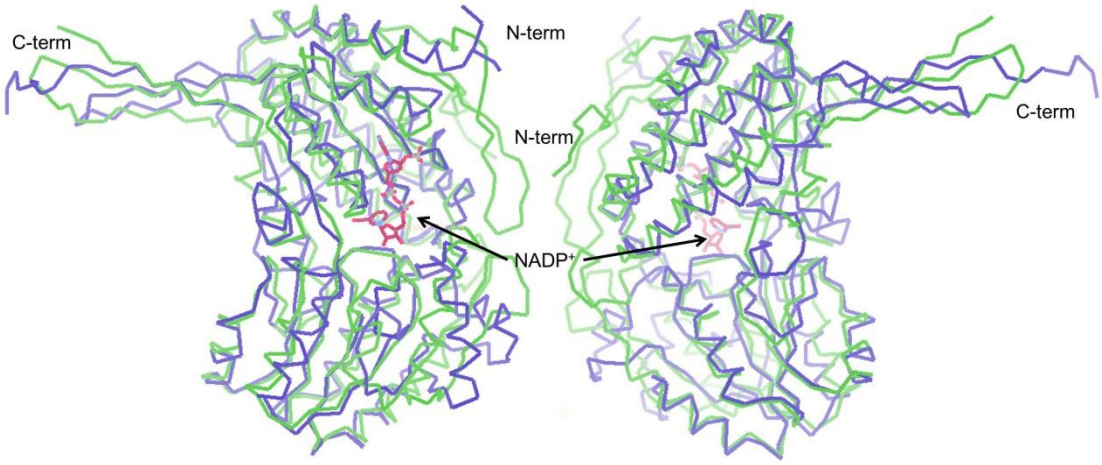


Figure 7-11: Cα traces of aligned A chains of AcAldDH (purple) and 3JZ4 (green); "front and back" views are shown. Visible termini are indicated and the $NADP^+$ molecule is shown in red.

The active site of the protein would appear to be arranged at the interface between the two domains of the protein and is positioned such that the active site is orientated away from the C-terminal dimerization face. Some clashes between the AcAldDH and modelled $NADP^+$ were observed, and are discussed during Section 7.5.5.

To further investigate the binding of substrates in AcAldDH, attempts were made to resolve the AcAldDH protein structure in the presence of substrates. Soaking experiments were attempted for the AcAldDH protein crystals that formed in the absence of substrates. Soaks for 30 s in cryo-protectant well-solution plus substrate mixtures, including acetyl-CoA (0.1 mM and 50 µM) or 1 mM acetaldehyde were tested. The acetaldehyde-soaked crystals did not diffract X-rays following soaking. The acetyl-CoA-soaked crystals did diffract X-rays, but no density for the substrate was observed in the processed data. Due to the lack of observable substrates following the soaking experiments, co-crystallisation experiments were attempted.

7.5 AcAldDH protein co-crystallisation with substrate-product mix

7.5.1 Crystallisation conditions

The purified His-tagged AcAldDH protein dialysed into 50 mM Tris-HCl pH 8.0, 150 mM NaCl, was screened with structure screen I & II, PGA, Morpheous, heavy and light twin pack and PACT screens (Molecular Dimensions, UK), at a protein concentration of 10.7 mg/ml. Two different concentrations of substrate-product mix were added to the protein solution immediately prior to the screens, identified as "low" (50 µM acetyl-CoA & 1.4 mM NAD^+) and "high" (0.5 mM acetyl-CoA & 5 mM NAD^+); this mixture of product and substrate was used in an attempt to obtain a structure with each of these molecules present to allow comparison of substrate/product binding. A 1:1 ratio (v/v) of protein solution: well solution (300 nl drop size) was used for the screening.

The best hit obtained was for the PACT screen in wells D2 & D3 (0.1 M MMT buffer pH 5.0 & 6.0, 25% (v/v) PEG 1500). To obtain the most suitable crystals for X-ray diffraction, optimisation of the conditions employed pH and PEG 1500 gradients.

Crystals producing the best diffraction data were grown at a protein concentration of 10.7 mg/ml diluted 1:1 with well solution (0.1 M MMT buffer pH 5.0, 21% and 23% (v/v) PEG 1500) with the "high" concentration of substrates. Crystals took approximately 4 weeks to appear at 16°C. The crystals were thin and plate-like. Stacks of crystals were often formed where new crystals nucleated on the side of existing crystals. Representative crystal forms are shown in Figure 7-12.

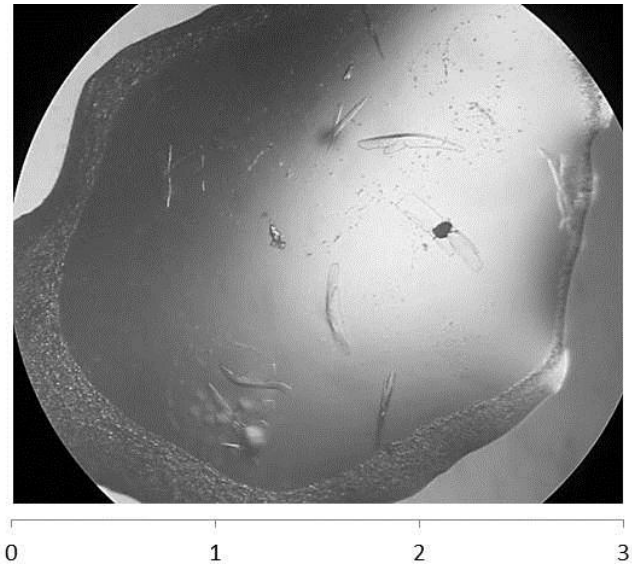


Figure 7-12: Representative crystal forms for AcAldDH co-crystallised with substrate-product mix (high) (scale-bar shown = mm).

Several cryo-protectants were tested with these crystals, including glycerol, ethylene glycol and formate over a range of concentrations. The crystals tested either dissolved in the cryo-protectant well-solution mix or did not diffract to any significant resolution when exposed to the X-ray beam. Freezing the crystals appeared to have a negative effect on X-ray diffraction; the room-temperature method was therefore employed in an attempt to improve resolution of the co-crystallised structure.

Cryo-crystallography, where crystals are exposed to the X-ray beam while in a steady stream of nitrogen gas, improves the lifetime of protein crystals (Garman and Owen 2006). Although the diffraction at room-temperature for the crystals was much improved compared to when frozen, the radiation damage experienced by the crystals was significantly increased (Nave 1995). This shortened the lifetime of the crystals and thus limited the completeness of the datasets obtained. Multiple partial datasets were collected and merged together to form the final dataset used.

7.5.2 Data collection and molecular replacement

Data were collected in-house using a MicroMax 007HF X-ray generator fitted with a Saturn 944+ detector (Rigaku, Japan); this used a copper K alpha rotating anode producing X-rays of wavelength 1.54060 Å. Data from 6 different crystals were used, and only crystals with unit cells within 0.5% for each axis were merged to ensure isomorphism. Images were separated by 0.5° of rotation. The merged dataset was processed using the HKL2000/0.98 software package. Data collection statistics are shown in the two following tables.

Crystal identity	Number of images	Unit cell dimensions
Crystal 1 (Rt 2001_2)	35	a = 93.054 Å, b = 109.543 Å, c = 204.149 Å, α=β=γ = 90.00°
Crystal 2 (Rt 2001_3)	20	a = 93.346 Å, b = 109.736 Å, c = 204.401 Å, α=β=γ = 90.00°
Crystal 3 (Rt 2001_4)	30	a = 93.181 Å, b = 109.479 Å, c = 204.252 Å, α=β=γ = 90.00°
Crystal 4 (Rt 2001_5)	24	a = 93.303 Å, b = 109.578 Å, c = 204.607 Å, α=β=γ = 90.00°
Crystal 5 (Rt 2001_6)	18	a = 93.200 Å, b = 109.581 Å, c = 204.306 Å, α=β=γ = 90.00°
Crystal 6 (Rt 3)	45	a = 93.109 Å, b = 109.538 Å, c = 204.094 Å, α=β=γ = 90.00°

Table 7-10: Crystal parameters for the AcAldDH (co-crystallised) individual datasets used for generation of the final merged dataset.

Parameter	Value
Average unit cell dimensions	a = 93.198 Å, b = 109.576 Å, c = 204.301 Å, α=β=γ = 90.00°
Space group	$P2_12_12_1$
R_{merge}	0.267 (0.466)
Completeness	78.0 (70.5)%
I/σ(I)	3.87 (2.26)
Multiplicity	2.5 (2.2)
Number of reflections observed	193392
Number of unique reflections	14328 (869)
Resolution	50-4.00 Å (4.07-4.00)

Table 7-11: AcAldDH (co-crystallised) data collection statistics (numbers in brackets for highest resolution bin).

The $p2_12_12_1$ AcAldDH structure without substrates was used as a starting model for the co-crystallised dataset. The model structure determined during Section 7.4 was rigid-body refined into the in-house data and the refinement process started as previously described.

7.5.3 Data refinement

Following the rigid-body refinement in Refmac5, the R factor was 0.3364 and R_{free} was 0.3355.

The model was subjected to several rounds of refinement and validation. Four ADP molecules and four chloride ions were added to regions of un-modelled density (one for each monomer), and the model refinement process continued until the solution was completed. Positive electron density was again observed at the terminal S_γ of Cys_{273} in all the AcAldDH monomers; the residue was replaced with an S-cysteinesulfinic acid during refinement as with the high-resolution structure described in Section 7.4.3. A dual conformation was observed for Glu_{287} in chain A; this was modelled with partial occupancy for both, and three water molecules were also added to the structure. Statistics for the refined final model are shown in Table 7-12. Following refinement, no difference map peaks ($F_o - F_c$) above σ level 5.0 were observed in the structure.

Up to 4 additional residues were distinguishable at the N-terminus of the AcAldDH co-crystallised structure compared to the high-resolution structure. Differences in the crystallisation conditions are likely to have affected the flexibility of these termini, allowing further residues to be resolved. The extensions of the N-termini suggest that the additional His-tags do not interact, but point out of the structure in different directions.

Parameter	Value
Number of reflections used	13488 (5% test set)
Number of protein atoms	13640
Number of non-solvent molecules	8
Number of solvent molecules	3
Overall mean B-factor ($Å^2$)	48.05
R	0.2264
R_{free}	0.3006
RMS Bond Length	0.0095 Å
RMS Bond Angle	1.5095°

Table 7-12: Final refinement statistics for the co-crystallised AcAldDH crystal structure with acetyl-CoA and NAD^+.

The structure was evaluated using MolProbity:

Parameter	Value
Ramachandran favoured	95.34% (1722/1807)
Ramachandran allowed	98.3% (1776/1807)
Ramachandran outliers	1.7% (31/1807)
MolProbity score	2.58 (98^{th} percentile)
Clashscore	20.73 (91^{st} percentile)

Table 7-13: AcAldDH (co-crystallised) MolProbity validation results.

The quality of the final structure is limited due to the relatively-low completeness of the data, caused by the requirement to merge several datasets together, as well as the limited resolution of the data used to produce the structure.

7.5.4 Overview of the co-crystallised structure

Although the resolution of the structure was poor compared to the structure in the absence of substrates, several regions of additional electron density were observed. The common ADP component of the substrates was successfully refined into the structure, but it was not possible to determine if the density was associated with NAD^+ or acetyl-CoA.

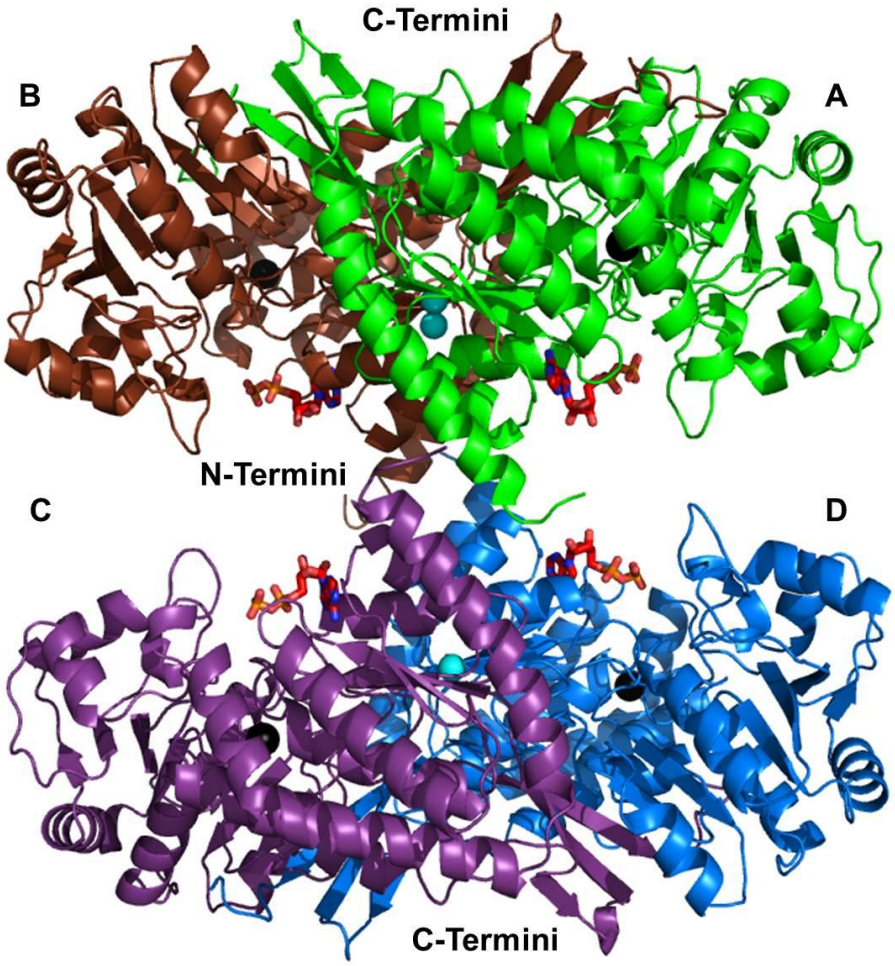


Figure 7-13: Cartoon overview diagram of the AcAldDH structure in the presence of NAD^+ and acetyl-CoA. Green = molecule A, brown = molecule B, purple = molecule C, and blue = molecule D. Black sphere = Cl^-, cyan spheres = water molecules, red molecules = ADP, and termini regions are indicated.

When the AcAldDH structure without substrate was superimposed onto this structure, each of the individual chains aligned well, but small structural differences were observed between molecules forming close dimeric interactions (i.e. molecule A and B or molecules C and D). However, the orientation of molecules forming the dimer of dimers appeared to have shifted more significantly, as shown in Figure 7-14. These differences may in part be due to differences in crystallisation conditions, but may also be due to the presence of acetyl-CoA and NAD^+. Structural alignment between the structures did not show RMSD values in excess of 0.8 Å between any of the chains, suggesting any differences were relatively small. Due to the limited resolution of the co-crystallised structure, meaningful interpretation of side-chain repositioning could not be made.

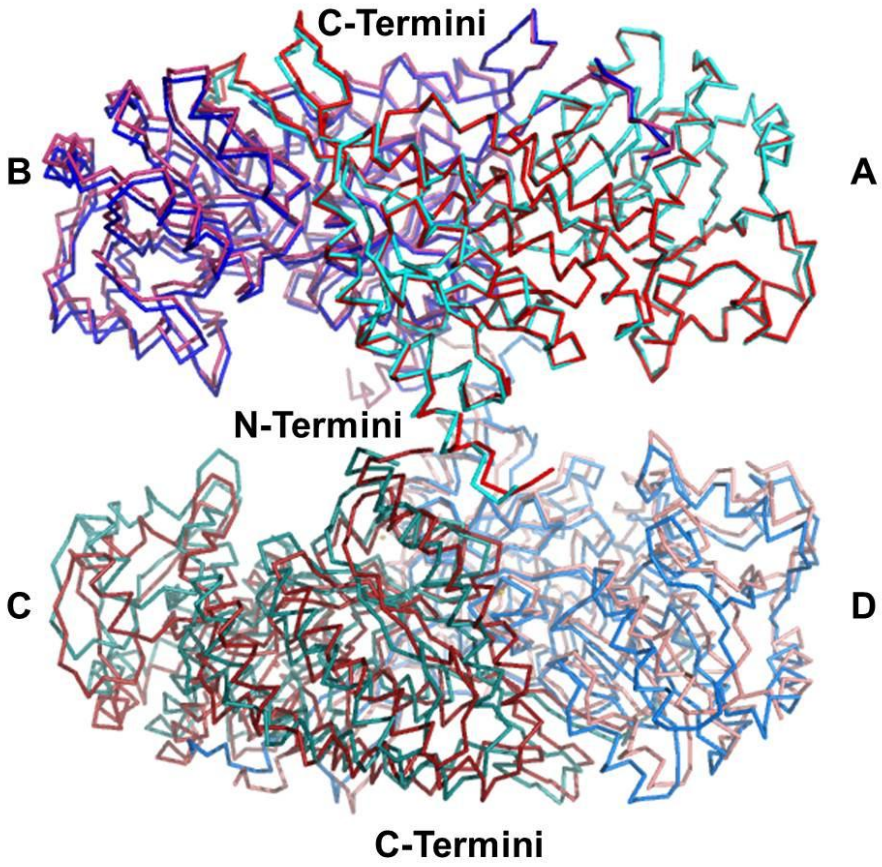


Figure 7-14: Cα traces of aligned AcAldDH tetramers. The A chain was used as reference for superimposition. The co-crystallised structure chains are shown in various shades of red, and the high-resolution structure chains are shown in shades of blue. Chain identities and terminal regions are indicated.

7.5.5 Regions of additional electron density

As shown in Figure 7-15, additional electron density was observed beside each of the AcAldDH chains proximal to His_{166}; this appeared to correspond to the shared ADP moiety that is present in both NAD^+ and acetyl-CoA (see Figure 7-19). The $NADP^+$ taken from the 3JZ4 structure shown in Section 7.4.7, occupied the same binding pocket when overlaid onto the co-crystallised AcAldDH structure; this implied that the ADP moiety was indeed likely to be interacting in this region.

The terminal phosphate oxygen of ADP is positioned between 10.6 and 11.6 Å from the Sγ of Cys_{273} for all the protein chains, allowing space for either the remaining portion of NAD^+ or acetyl-CoA to fit. Unfortunately, sufficient density was not present beyond that associated with the ADP to allow the substrate/product identity to be distinguished.

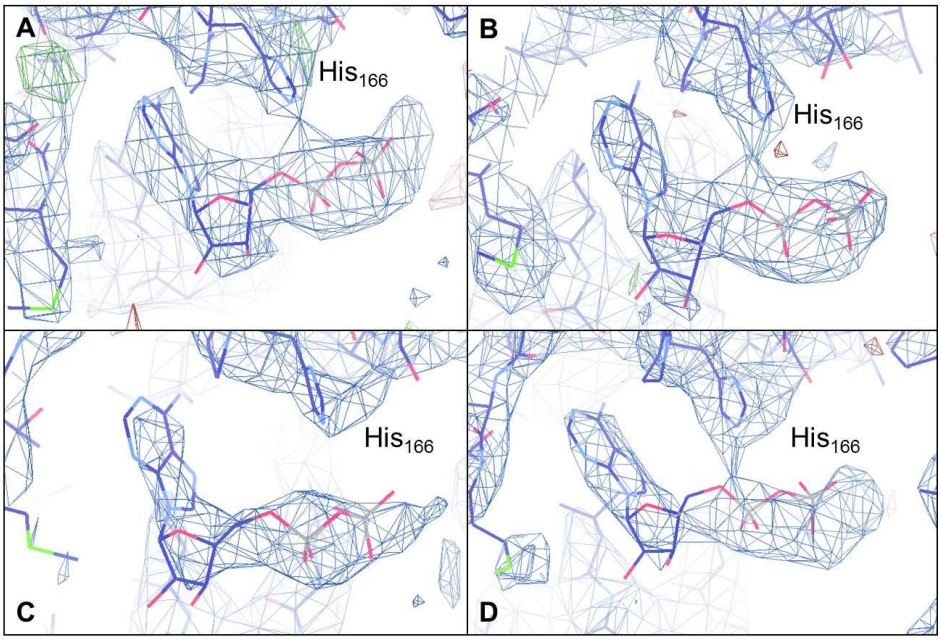


Figure 7-15 A-D: Electron density maps for the 4 ADP molecules in the co-crystallised AcAldDH structure (A-D = chain id in structure). Blue map = $2F_o$ - F_c (where F_o = $F_{observed}$, F_c = $F_{calculated}$, and F = structural factor), σ level = 1.09; negative (red) and positive (green) difference map (F_o – F_c), σ level = 3.0. Molecules are shown in stick form, with carbons in purple.

The presence of an ADP in this region of the protein confirms the orientation of the active-site cleft as facing inwards towards the N-terminal faces of the monomers (Figure 7-13). The observed tetrameric assembly may offer some protection to the active-site cysteines by orientating all the active sites into the centre of the multimer rather than their being exposed on the surface of the protein.

As shown in Figure 7-16, the positioning of the ADP appeared significantly different from that portion of $NADP^+$ in 3JZ4; this may be expected due to the presence of clashing residues when this $NADP^+$ was superimposed onto the AcAldDH structure. Significant steric clashes were observed between the 2' additional phosphate of $NADP^+$ and a loop containing residues His_{166} – Ser_{168}. Given that AcAldDH would appear to bind acetyl-CoA/Co-ASH in this region, it is likely that some differences in substrate orientation may exist between the two structures. The orientation observed in the AcAldDH structure would position the 2' and 3' phosphate on the ADP ribose in $NADP^+$ and acetyl-CoA, respectively, away from the active site alleviating any steric clashes observed. This may explain how CoA-SH and acetyl-CoA are able to bind at the same site as NAD^+ in the AcAldDH protein. The adenosine ring observed here is also in a different position from the modelled $NADP^+$; this positioning avoids the clash predicted between this ring and Thr_{205}.

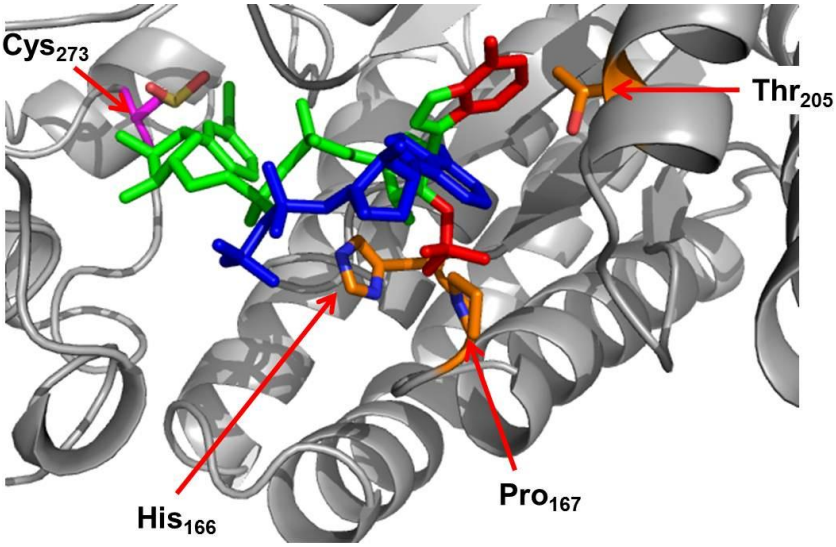


Figure 7-16: Cartoon diagram of the co-crystallised AcAldDH active-site region in the presence of NAD^+ and acetyl-CoA with $NADP^+$ from 3JZ4 overlaid. Grey = molecule A, Cys_{273} is shown with purple carbon atoms. Residues clashing with $NADP^+$ are shown with orange carbon atoms. Green molecule = superimposed $NADP^+$, red regions of $NADP^+$ indicate clashes, blue = ADP moiety visible in the structure (ligands shown in stick form).

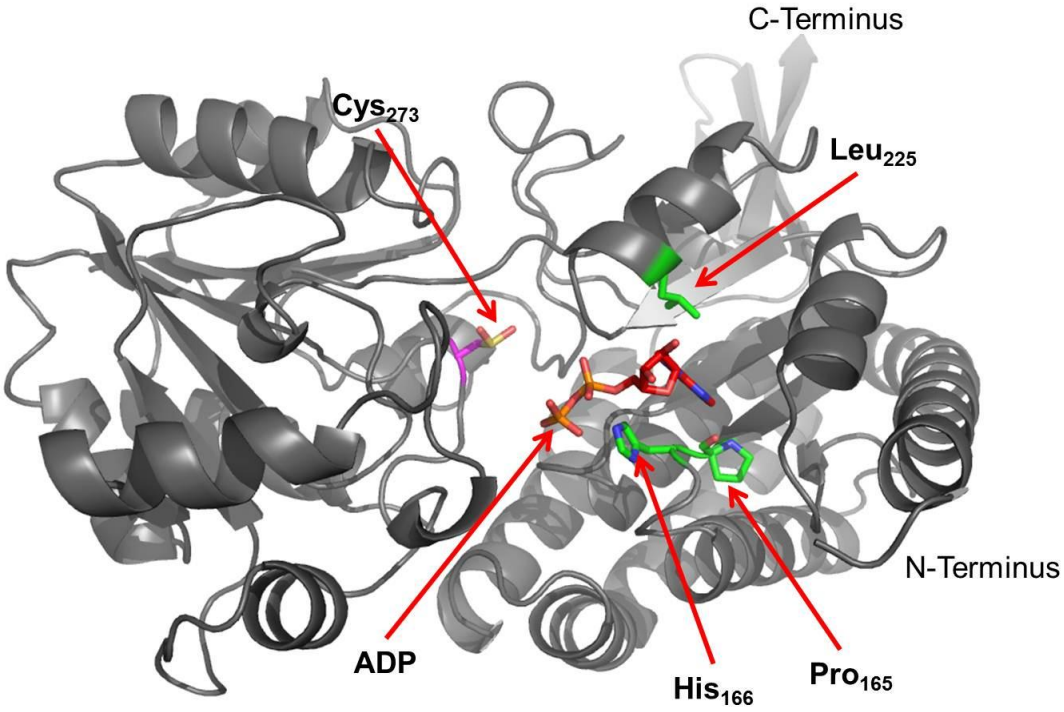


Figure 7-17: Cartoon overview diagram of the AcAldDH structure (molecule A) in the presence of NAD^+ and acetyl-CoA. Protein termini, Cys_{273} (purple), residues interacting with ADP in all chains (shown in green) and the ADP moiety (stick form with carbons shown in red) are indicated.

Detailed analysis of ligand binding could not be carried out due to the limited resolution of the cofactor-bound structure. Uncertainty surrounding the precise conformation of interacting residues was of particular concern. Given the substantial involvement of

water molecules in substrate binding observed for 3JZ4 (Langendorf et al. 2010), the low-resolution structure generated here can only provide limited information. According to CONTACT analysis (CCP4i), Pro_{165}, His_{166} and Leu_{225} are interacting with ADP in all chains as shown in Figure 7-17.

7.6 Discussion

The AcAldDH protein from *G. thermoglucosidasius* has been structurally resolved to 2.1 Å resolution in the absence of substrates; a less complete, lower resolution structure (4.0 Å) has also been obtained in the presence of NAD^+ and acetyl-CoA. As discussed in the Introduction to this thesis, the conserved residues in the aldDH domain of ADHE are conserved in the AcAldDH protein; the structure also aligns well with the 3K9D structure used for molecular replacement, showing these proteins to be of a similar topology. The conserved catalytic cysteine residue (Cys_{273}) identified through sequence alignments, appeared to have been oxidised due to radiation damage during the data collection process. This residue is positioned proximal to the likely position of the nicotinamide ring of NADH as required to perform catalysis.

In contrast to the aldDH domain of ADHE, the AcAldDH protein did not appear to show any oxygen sensitivity during biochemical characterisation. As the catalytic mechanism for the two proteins is likely to be the same, it is not clear why inactivation caused by the hypothesised oxidation of the catalytic cysteine (observed during Chapter 3) did not affect the AcAldDH protein. Alignment of the aldDH domain model (Chapter 5) with the AcAldDH structure obtained here (Figure 7-18), suggests no significant difference in key-residue positioning or interactions in the two proteins. The aldDH structure is only a homology model, and without a high-resolution structure of the aldDH domain it is impossible to accurately rationalise the differences in stability observed for the two enzymes. One possibility is that interactions formed between the ADH and aldDH domains in ADHE, may leave the aldDH domain more vulnerable to oxidation; the AcAldDH protein that is produced in isolation may be free to form tetrameric assemblies which may protect the catalytic cysteine. The catalytic clefts of the proteins appear to be reasonably open, which may leave this residue vulnerable if inter-protein interactions were to be disrupted.

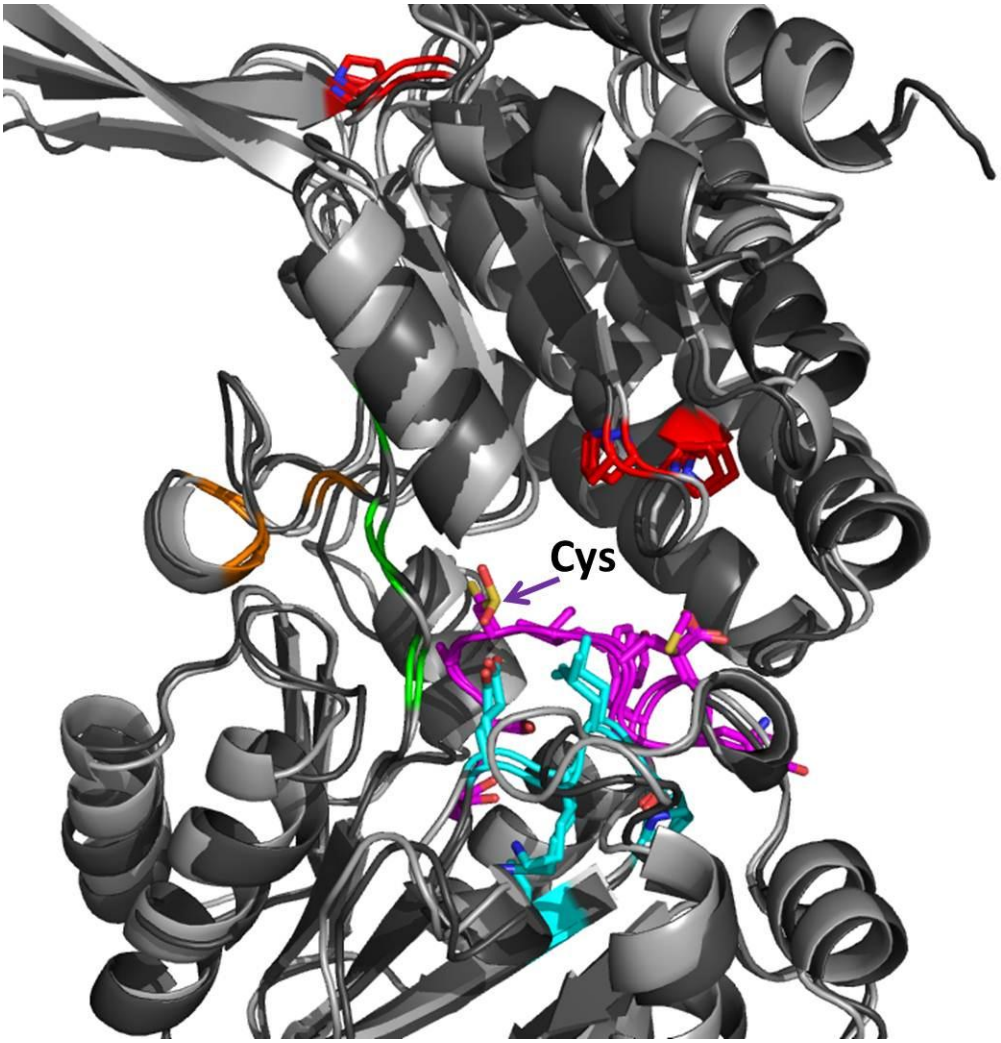


Figure 7-18: Cartoon diagram of an AcAldDH monomer (dark grey) superimposed on the aldDH model (light grey) created during Chapter 5. Conserved residues described in the Introduction to this thesis are indicated. Red = PXGXXXXXPXXXP motif, green = GXGXG motif, purple = catalytic cysteine containing motif, orange = GXGXXG motif and light blue = EKLSPXL motif. The conserved cysteine is labelled.

A "bi-uni-uni-uni ping-pong" scheme has been suggested for other acetylating aldDH enzyme mechanisms, where there is an ordered binding of substrates. This scheme has been investigated in *E. coli* and *Clostridium kluyveri* in the "reverse" direction to the assays described here, where acetaldehyde and NAD^+ bind to the enzyme first, and NADH is then released prior to CoA-SH binding and subsequent acetyl-CoA release (Shone and Fromm 1981; Smith and Kaplan 1980). The catalytic cysteine conserved between various aldDH proteins (described in the Introduction to this thesis) has been suggested to play a key role in the formation of an acetyl-S-enzyme intermediate during this reaction. The observation of an acetate close to the catalytic cysteine in the high-resolution structure (as discussed in Section 7.4.5), supports the prediction of this intermediate forming during the reaction mechanism. Smith and Kaplan (1980) hypothesise that the NAD^+ and CoA-SH molecules share a binding site on the protein,

leading to the independent binding during the reaction mechanism. Figure 7-19 shows the significant homology around the adenosine diphosphate regions of NAD^+ and CoA-SH. It can be envisaged that, due to the similar composition of this region, the two molecules may both be stabilised by the same interacting residues in the active site region. Variation in stabilising residues may occur around the non-homologous regions of these substrates.

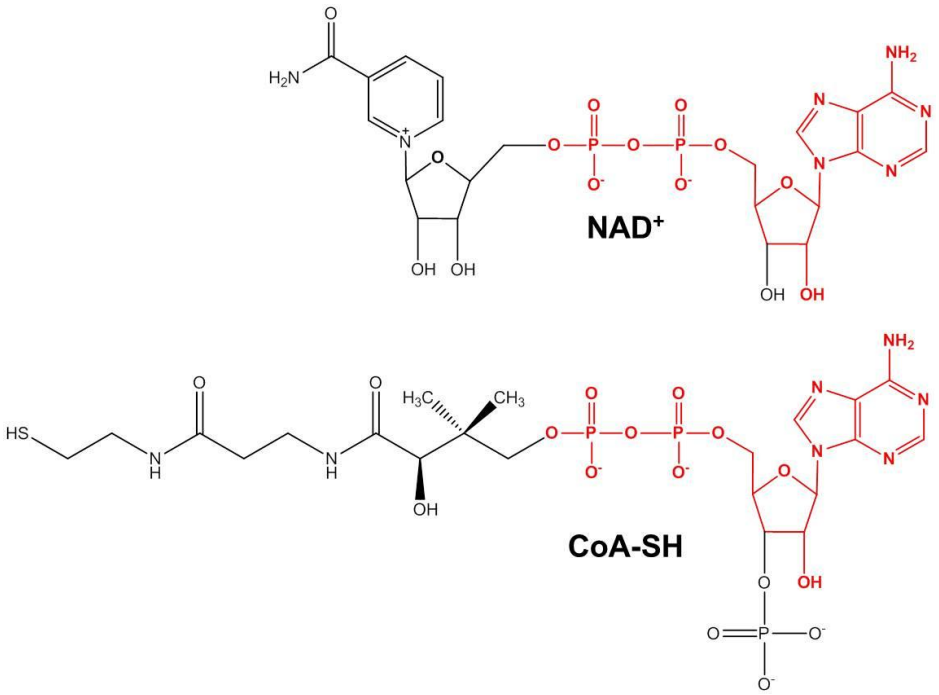


Figure 7-19: Chemical structure of NAD^+ and CoA-SH, with homologous "ADP" regions indicated in red.

Product inhibition was observed for the AcAldDH protein with respect to CoA-SH and acetaldehyde (Chapter 6). A shared binding site may explain this observation where the inefficient release of CoA-SH impedes NADH binding and hence limits catalysis. Inhibition by acetaldehyde cannot be rationalised by this mechanism, but the accumulation of this highly reactive product could be envisaged to have a negative influence on protein stability/activity. The DmpF protein, which is an acetylating aldDH from *Pseudomonas* sp. CF600, has been suggested to possess a shared binding site in a Rossmann fold for CoA-SH and NAD^+, as informed by an X-ray crystal structure and deuterium-exchange experiments (Lei et al. 2008).

In contrast to AcAldDH, the aldDH domain of ADHE did not display any product inhibition. Possible differences in the binding of the substrates/products may be the cause of the differences observed; the amino acid sequence of AcAldDH is 43% identical and 59% similar to that of the aldDH domain of ADHE, and thus significant

differences in interacting residues outside the common aldDH residues may occur. The presence of the ADH domain may also act as an influence over efficient product release. The difficulty in resolving the aldDH domain of ADHE independently of the ADH domain prevented examination of these potential factors.

Although diffraction data for AcAldDH crystals in the presence of acetyl-CoA and NAD^+ were obtained, as well as density indicative of portions of these substrate/products being visible, poor diffraction and the short lifetime of the AcAldDH crystals in the X-ray beam at room-temperature prevented full resolution of the hypothesised shared binding site. Several different crystals were used to generate the model and differences in occupancy of the molecules may have affected the density observed for the substrates. The additional density observed in the structure did appear close to the position of the modelled $NADP^+$ diphosphate (from 3JZ4) and the conserved ADP portion fitted well into this density. The conformation of the ADP refined into the structure was significantly different from the conformation of this moiety in the modelled $NADP^+$ from 3JZ4; it is possible that substrate orientation in AcAldDH may be different, avoiding clashes with active site residues, although it should be acknowledged that the limited quality of the co-crystallised structure limits the reliability of any conclusions drawn. During the refinement, no other regions of additional density were observed corresponding to portions of NAD^+ or acetyl-CoA molecules. This supports the prediction of a single binding site for these two substrates. Unfortunately, no similar protein structures were available with acetyl-CoA or CoA-SH bound, so the binding of this substrate could not be validated though modelling work.

Residues close to the additional 3' phosphate group of the ribose ring may present a suitable target for manipulation for optimisation of catalysis in the acetaldehyde generating direction. CoA-SH must be released prior to NADH binding, so decreasing the affinity of AcAldDH for CoA-SH may overcome the product inhibition observed previously. However, disruption of such binding interactions may significantly disrupt catalysis carried out by the enzyme. The orientation of ADP in the co-crystallised AcAldDH structure also suggests that this 3' phosphate region may not be directly interacting with the protein during substrate/product binding. Residues interacting with the non-homologous regions of the substrates may also provide targets for manipulation, but regrettably these interactions were not observed here.

Optimisation of co-crystallisation conditions with each of these molecules separately and of data collection conditions, will be required to allow this predication of a shared

binding site to be further investigated. Of particular interest is the identification of any residues that solely interact with acetyl-CoA/Co-ASH, rather than NAD^+, informing site-directed mutagenesis attempts to overcome product inhibition.

A particular focus will be the development of freezing conditions in an attempt to allow exposure to high-energy synchrotron radiation, increasing the resolution of the substrate/product bound structures. This may also allow any structural changes required in the protein during catalysis to be identified.

Another approach used in attempts to alleviate the product inhibition was to fuse the AcAldDH protein to the ADH protein Fragment 11, to facilitate substrate channelling between the two proteins (Chapter 8). Structural models of the fusion proteins created in Chapter 8 were not produced due to the uncertainty of interactions that may form between the two fused domains (particularly for the more stable Fusion 1 protein); interactions may be non-native due to the artificial fusion of the two domains, and therefore more difficult to accurately predict compared to the physiologically-relevant ADHE model described in Chapter 5.

In conclusion, the AcAldDH protein of *G. thermoglucosidasius* has been structurally resolved to 2.1 Å resolution by X-ray crystallography. Active site residues have been identified; however, due to the limited resolution of the substrate/product bound structure, accurate identification of potential target residues for manipulation of enzymatic activity has proven elusive. Confirmation of a likely shared binding site for NAD^+ and CoA-SH suggests alleviating product inhibition through point mutations may not be feasible, due to shared interactions between the protein and substrates.

8 CREATION OF NOVEL ARTIFICIAL ADHE PROTEINS

8.1 Introduction

The aim of this section of the project was to produce a catalytically-active artificial ADHE enzyme, derived from the Fragment 11 and AcAldDH proteins expressed and characterised previously. The effects of such fusions on the two domains in terms of stability and activity were monitored. The kinetic efficiency of these enzymes expressed in isolation appeared to be limited by certain properties of the proteins. The AcAldDH appeared to be subject to product inhibition, potentially due to the inefficient release of the products of the catalysed reaction. Limitations in catalytic efficiency were observed for the ADH domain of the protein, indicated by a relatively high K_m for acetaldehyde. *In vitro* experiments showed co-incubation of the two proteins to be insufficient to allow conversion of acetyl-CoA to ethanol. Either of these described limitations could provide an explanation as to why the two step catalysis was not observed. Channelling of acetaldehyde may be induced by positioning the constituent proteins (and preferably the active sites) close together by expression with a physical link between them.

It was considered of academic and industrial interest to observe the effect of expressing these two proteins as a single protein unit. The proteins produced were evaluated in terms of *in vitro* catalytic properties and also *in vivo* in terms of modulation of ethanol production within *Geobacillus* strains. The ability of the soluble Fragment 11 protein to induce proper folding of the EutE aldDH protein, when expressed as a protein fusion, was also tested.

Fusions of the AcAldDH and Fragment 11 proteins were made in two orientations. Fusion 1 consisted of an N-terminal Fragment 11 and a C-terminal AcAldDH. Fusions 2-4 were arranged in the opposite orientation as seen for the native ADHE protein. A schematic diagram of the protein Fusions is shown in Figure 8-1.

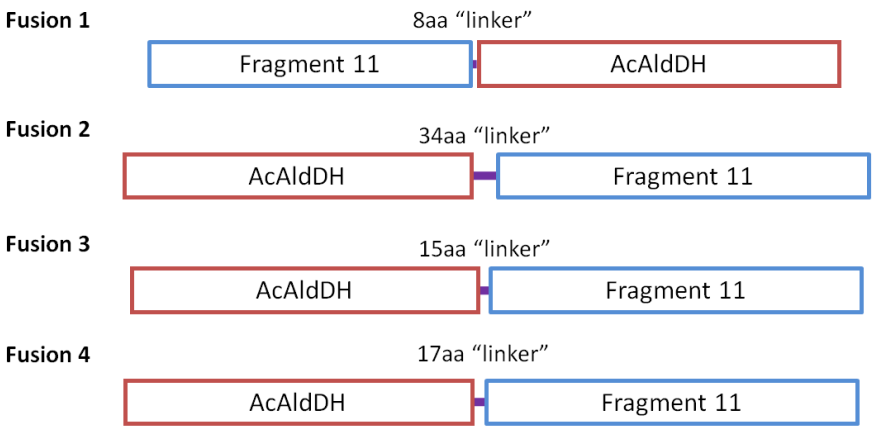


Figure 8-1: Schematic diagram of the un-tagged Fusion proteins created during this chapter of the project. "Linker" lengths shown are based on the amino acid sequence between structured domains of the proteins (determined from the crystal structures of the two domains) plus any additional sequence added through the vector used for the fusions.

The variable DNA sequence of the linker region between the two domains for Fusion proteins 2-4 are shown in Figure 8-21.

8.2 Materials and methods

DNA templates for the PCR amplification of the genes described were either TM242 genomic DNA (50 ng/µl), or the previously-described plasmid constructs. These were carried out according to the method described in Section 2.3.2. The DNA primer combinations used to amplify the genes are shown in the following Tables. Primer DNA sequences are reported in Appendix 1. Primers incorporated additional bases where required to keep the two proteins to be fused and any additional tags in frame. Reverse primers for the N-terminal gene sequences did not include a stop codon to allow read through. Reverse primers for the C-terminal gene sequences incorporated a stop codon immediately prior to the 3' restriction site.

The PCR products for the generation of fusion proteins were A-tailed, ligated into the pGEM®-T easy Vector, ethanol-precipitated and transformed into JM109 cells for blue/white screening. Successful cloning of the fragments was confirmed by DNA sequencing. The fragments in pGEM®-T easy were then digested with the appropriate restriction enzymes, gel purified and ligated into the previously-digested expression vectors. Ligations were transformed into JM109 cells, which were then subject to PCR colony screening. Positive constructs from the PCR screen were screened by restriction digestion, and were then sent for DNA sequencing with vector-specific and internal primers for the appropriate gene orientation. Bacterial strains transformed with

the various expression vector constructs were screened using appropriate antibiotic selection, using antibiotic resistance markers present in the vectors.

8.2.1 Fusion 1 cloning

The pLM303-*mbp-eutE* construct produced in Chapter 6 was used as the starting point of the Fusion 1 protein construction. The maltose-binding protein was replaced by Fragment 11 to produce the pLM303-*frag11-eutE* construct, and the *eutE* gene was replaced with *AcAldDH* to produce Fusion 1. For expression in *Geobacillus* the internal SacI site was disrupted using site-directed mutagenesis incorporating an additional codon GGT (G). No His-tag was present on the Fusion 1 protein expressed from pUCG18-pLDH. Primers used during this process are shown in Table 8-1.

Cloning stage	Restriction sites	Forward primer	Reverse primer
Fragment 11 cloning for EutE fusion	NcoI/SacI	Frag11 for eutE fusion F1	Frag11 for eutE fusion R1
AcAldDH replacement of EutE in fusion	SacI/XhoI	ACaldDH fusion Fwd1	ACaldDH pet28 rev1
Removal of SacI site (mutagenesis)	n/a	11-AcaldDH NO SACI Fwd:	11-AcaldDH NO SACI Rev
Amplification of Fusion 1 for expression in pUCG18-pLDH vector	XbaI/SacI	GB 459-869aa Fwd1	pUC F1 Rev1

Table 8-1: PCR primers used for construction of the Fusion 1 protein.

8.2.2 Fusions 2-4 cloning

The *fragment 11* DNA sequence was cloned into pET28a. The a*cAldDH* (without a stop codon) was then introduced at the N-terminus of this gene to allow read through, creating the N-terminal His-tagged Fusion 2 protein. The Fusion 2 protein without the N-terminal His-tag was selected for expression in *Geobacillus*. Primers used during this process are shown in Table 8-2.

Cloning stage	Restriction sites	Forward primer	Reverse primer
Amplification of Fragment 11 for Fusion 2	HindIII/XhoI	Frag 11 28a Fusion 2 F1	Frag11 pET28a R
Amplification of AcAldDH for Fusion 2	NheI/HindIII	ACaldDH pet28 F1	AcAldDH 28a Fusion2 R1
Amplification of Fusion 2 for pUCG18-pLDH vector	XbaI/SacI	pUC F2 Fwd1	pUC F2 Rev1

Table 8-2: PCR primers used for construction of the Fusion 2 protein.

Proteolysis was observed within the linker region of the recombinantly-expressed Fusion 2 protein. Truncation of the AcAldDH gene sequence with a slight extension of the Fragment 11 sequence was used to create Fusion 3. Codons for two additional amino acids (GS) were introduced between the two domains of Fusion 3 using site-directed mutagenesis, disrupting the HindIII restriction site. The resulting protein was known as Fusion 4. Primers used during this process are shown in Table 8-3.

Cloning stage	Restriction sites	Forward primer	Reverse primer
Amplification of AcAldDH for Fusion 3	NheI/HindIII	ACaldDH pet28 F1	AcAldDH pet28a R1 Native link
Amplification of Fragment 11 for Fusion 3	HindIII/XhoI	ADH native link F1	Frag11pET28aR
Mutagenesis of Fusion 3 to create Fusion 4	n/a	Fus4 Fwd primer	Fus4 Rev primer

Table 8-3: PCR primers used for construction of the protein Fusions 3 & 4.

8.2.3 TM400 Fusion 1 vs ADHE fermentation run

To evaluate the Fusion 1 protein *in vivo,* TM400-pUCG18-pLDH-*adhE* and *fusion 1* strains were grown under fermentative conditions using Biostat CT-DCU (1 L vessel) fermenter systems (Sartorius), following TMO Renewables standard protocols. The methods used are summarised in the results section of this chapter. The fermentation runs were carried out in conjunction with the fermentation team at TMO Renewables. The TMO analytical team determined metabolite concentrations using HPLC analysis, following standard protocols.

8.3 Results

The orientation of the component protein domains appeared to have a significant effect on the stability of the Fusion proteins. Protein expression and characterisation results for the two orientations will be reported separately.

8.3.1 Fusion 1

8.3.1.1 Fusion 1 cloning

Restriction digests of the final expression constructs are shown in Figure 8-2. The pLM303-*fusion 1* construct was used for expression in *E. coli* of the His-tagged Fusion 1 protein. Later experiments expressing the protein in *Geobacillus* strains used the un-tagged pUCG18-pLDH-*fusion 1* construct.

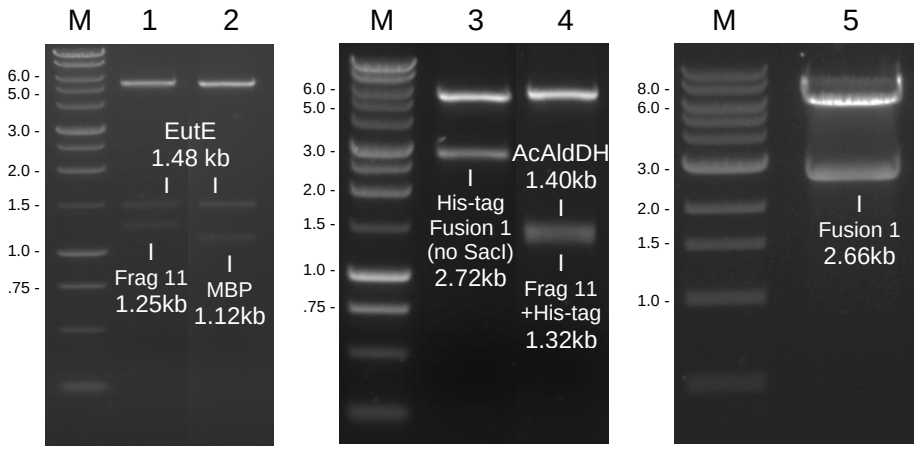


Figure 8-2: Agarose gel electrophoresis of restriction-digested constructs. The predicted sizes of the various DNA fragments are shown over the respective lane for each digest. M = DNA markers sizes given in kb, 1 = pLM303-*frag11-eutE*, 2 = pLM303-*mbp-eutE* (both triple digested SacI/NcoI/XhoI), 3 = pLM303-*fusion 1* (no SacI site), 4 = pLM303-*fusion 1* (with SacI site) (both triple digested XbaI/SacI//XhoI) & 5 = pUCG18-pLDH-*fusion 1* (digested with XbaI/SacI).

8.3.1.2 Fusion 1 (and Frag11-EutE) protein expression and characterisation

The purified pLM303-*frag 11-eutE* & pLM303-*fusion 1* plasmid constructs were transformed into the *E. coli* BL21 (DE3) protein expression strain. Protein expression trials were carried out using 4.5 h of protein expression conditions.

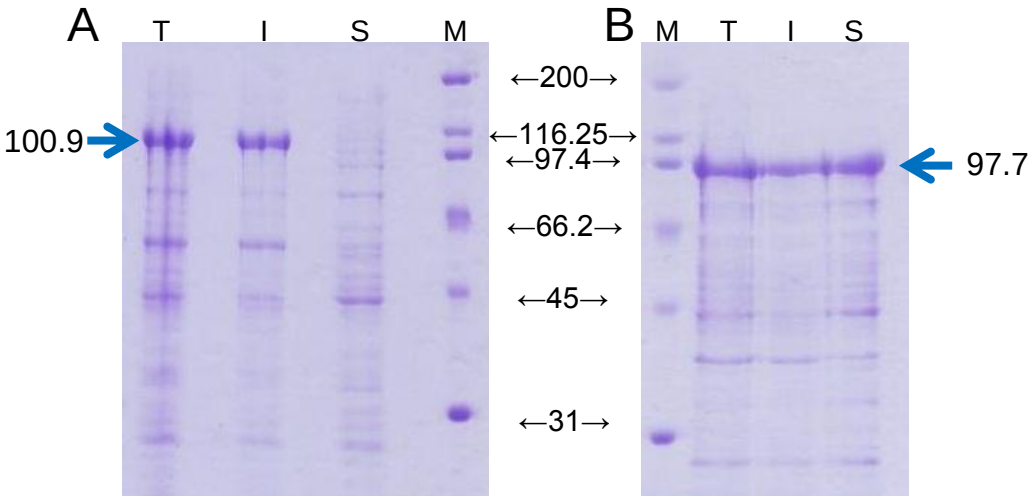


Figure 8-3: SDS-PAGE analysis of BL21 Fusion protein expression trial. M = markers (M_r/1000), T = total, I = insoluble & S = soluble. A = Frag 11-EutE (M_r = 100,924), B = Fusion 1 (M_r = 97,705). Predicted His-tagged-protein bands are highlighted with a blue arrow, and the expected size (M_r/1000) of the protein is shown.

The Frag11-EutE protein was produced insolubly under the expression conditions tested, whereas the Fusion 1 protein was expressed in the soluble cell extract. Fusion 1 produced under the described expression conditions was purified using metal

affinity chromatography, as shown in Figure 8-4 (column charged with Ni^{2+} according to the method outlined in Section 2.9).

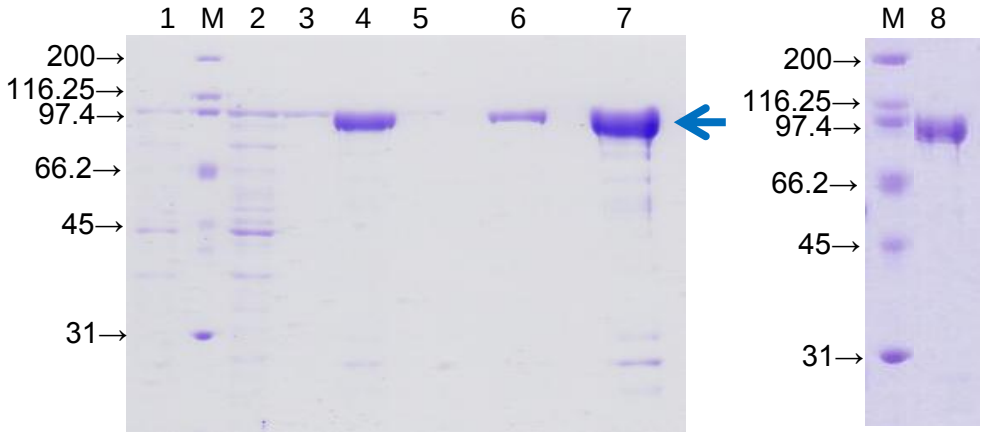


Figure 8-4: SDS-PAGE analysis of Fusion 1 metal affinity chromatography samples. 1 = flow through, M = markers (M_r/1000), 2 = 0%, 3 = 1%, 4 = 2 x 15% elution step pool, 5 = 100%, 6 & 7 = 15% elution 0.5 and 2 x sample concentration & 8 = 15% pooled sample 8 days post purification (% values = %HIS-ELUTE buffer diluted in HIS-BIND buffer). Predicted His-tagged-Fusion 1 band (M_r/1000 = 97.7) is highlighted with a blue arrow.

The pooled 15% HIS-ELUTE sample containing the highest ADH and aldDH activities was dialysed overnight into 50 mM Tris-HCl pH 8.0, 150 mM NaCl. Post dialysis, some protein had precipitated in the tubing, and therefore the sample was filtered using a 0.44 µm syringe filter device (Millipore) prior to assays. A 45% loss in protein concentration was observed post dialysis. Although a significant loss of protein activity was also observed at this stage, significant ADH and aldDH activities remained present in the sample. Assessment of the solution by SDS-PAGE following 8 days incubation at 4°C (lane 8 in Figure 8-4) showed the fusion protein had not been degraded over this time.

Assays were undertaken to determine the kinetic parameters for the two domains of this protein. As in Chapter 6, rapid curvature of the progress curves was observed for the AcAldDH assays. Assays of this domain were carried out as described previously.

Activity (Substrate)	K_m (mM)	Standard Error	V_{max} (U mg^{-1})	Standard Error
AcAldDH: (acetyl-CoA)	0.032	+/- 0.001	30.2	+/- 1.0
AcAldDH: (NADH)	0.057	+/- 0.002	28.2	+/- 1.0
Frag 11: (acetaldehyde)	151.0	+/- 5.5	145.9	+/- 5.9
Frag 11: (NADH)	0.131	+/- 0.004	110.6	+/- 4.3

Table 8-4: Summary of kinetic parameters determined for the Fusion 1 protein. V_{max} reported has been adjusted for saturation with the fixed substrate.

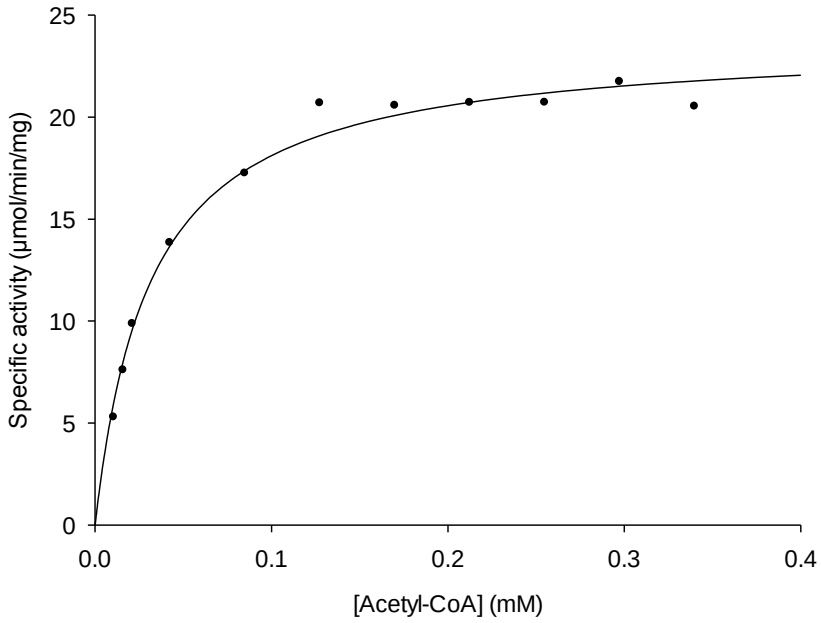


Figure 8-5: Michaelis-Menten plot of aldDH enzymic activity (U mg^{-1} of protein) versus concentration of acetyl-CoA (mM) at a fixed concentration of NADH (0.22 mM) for the purified Fusion 1 protein.

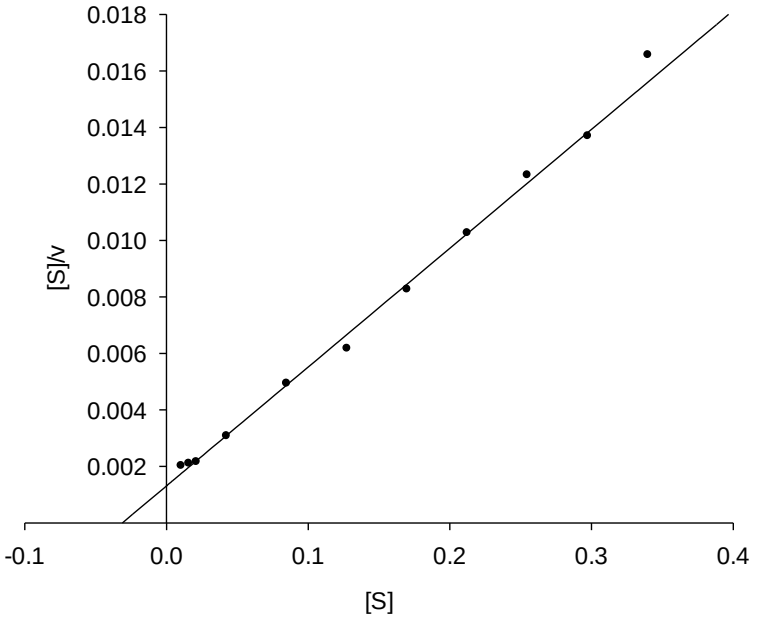


Figure 8-6: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of aldDH activity (U mg^{-1}) with respect to concentration of acetyl-CoA (mM).

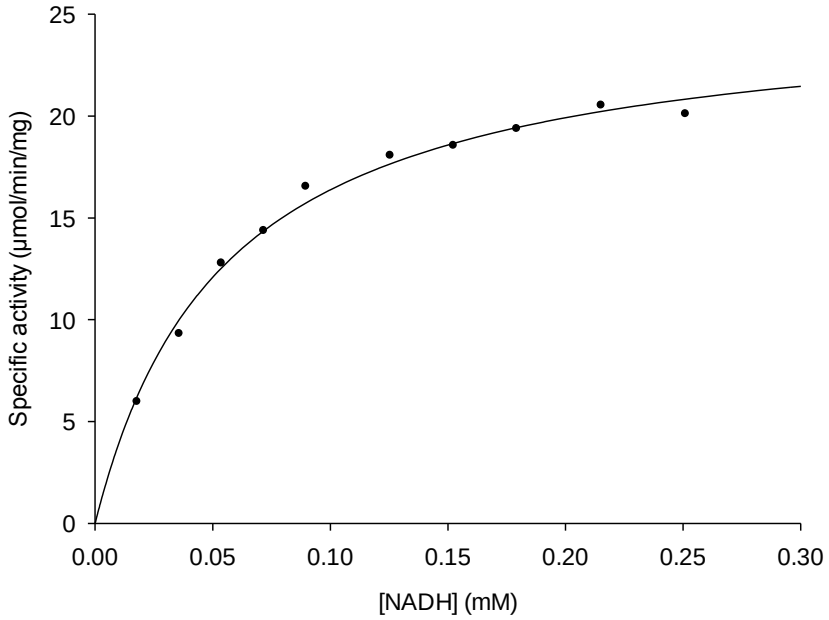


Figure 8-7: Michaelis-Menten plot of aldDH enzymic activity (U mg^{-1} of protein) versus concentration of NADH (mM) at a fixed concentration of acetyl-CoA (0.34 mM) for the purified Fusion 1 protein.

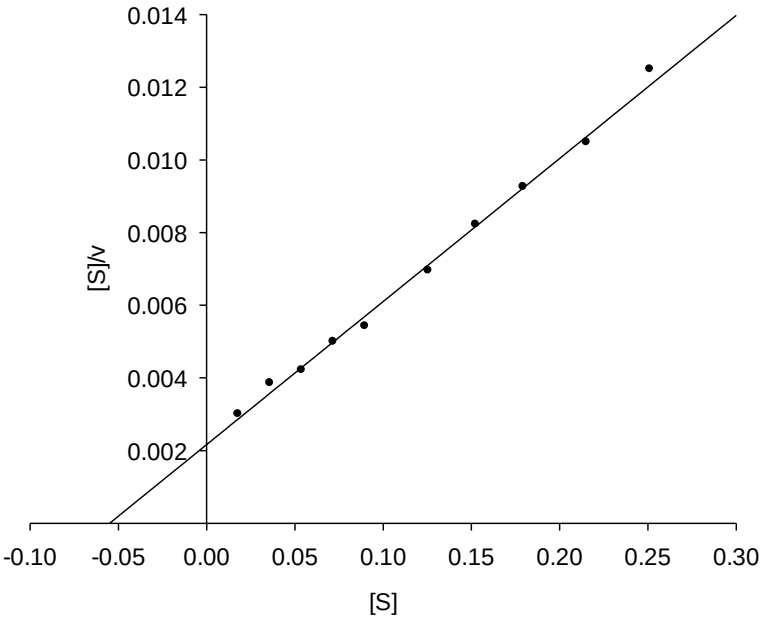


Figure 8-8: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of aldDH activity (U mg^{-1}) with respect to concentration of NADH (mM).

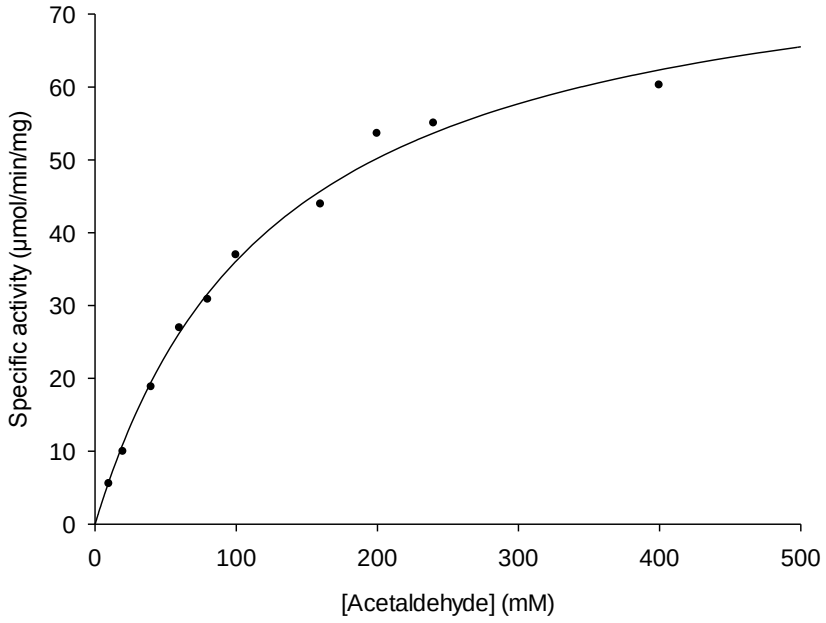


Figure 8-9: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) versus concentration of acetaldehyde (mM) at a fixed concentration of NADH (0.21 mM) for the purified Fusion 1 protein.

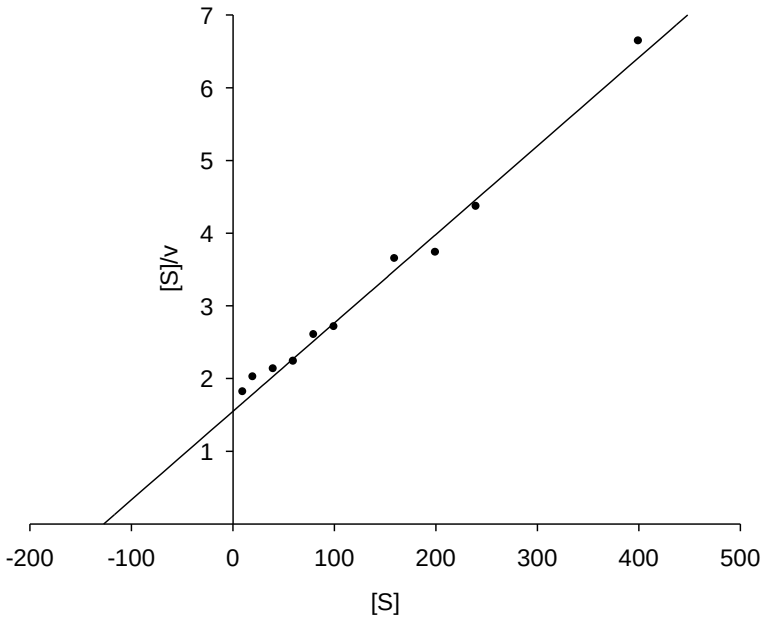


Figure 8-10: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of acetaldehyde (mM).

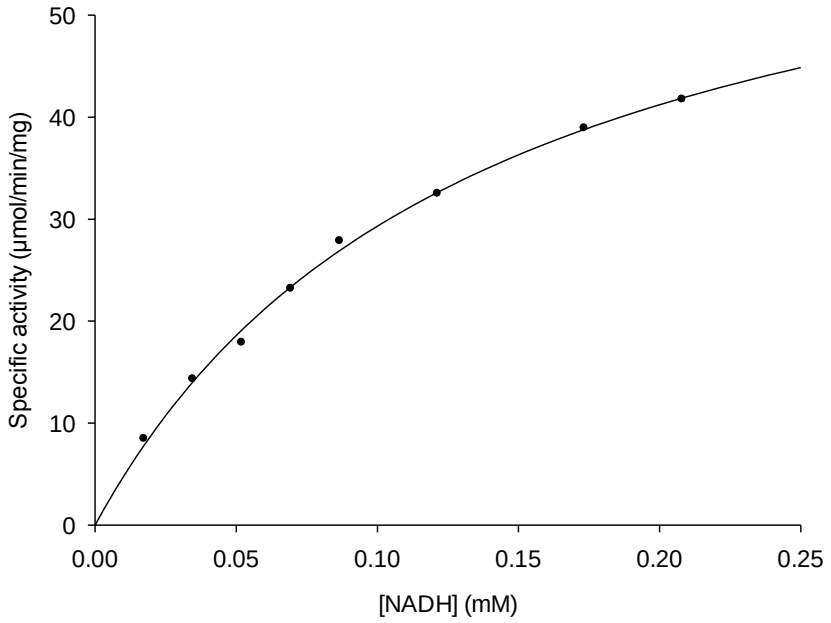


Figure 8-11: Michaelis-Menten plot of ADH enzymic activity (U mg^{-1} of protein) versus concentration of NADH (mM) at a fixed concentration of acetaldehyde (240 mM) for the purified Fusion 1 protein.

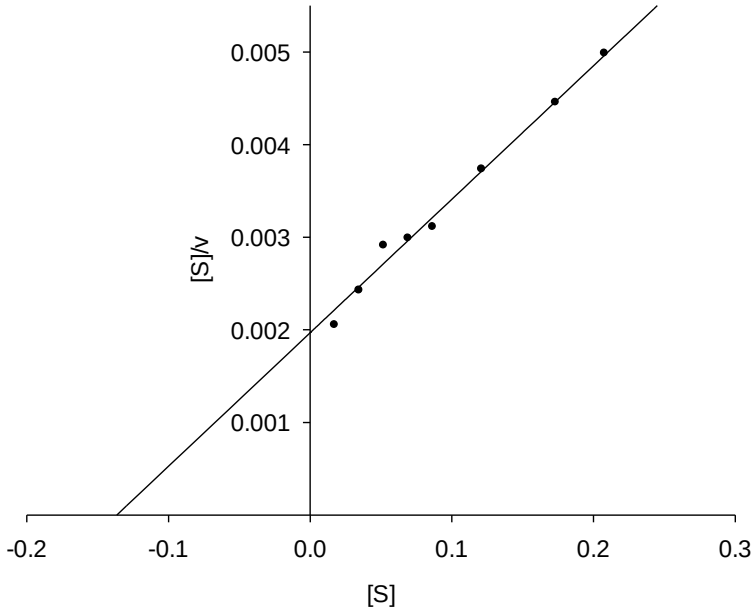


Figure 8-12: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of ADH activity (U mg^{-1}) with respect to concentration of NADH (mM).

Assays to determine whether the Fusion protein was capable of catalysing the conversion of acetyl-CoA to ethanol were carried out as described in Section 2.13.4. An average of three assays gave a ratio of 1.2:1 NADH converted to CoA-SH released. If the ADH domain of the protein was able to utilise the acetaldehyde produced by the AcAldDH domain, two NADHs would have been converted per CoA-SH released.

8.3.1.3 Fusion 1: optimum temperature and thermostability

Assays to determine the temperature optimum were carried out for the purified Fusion 1 protein as described in Section 2.13.5.

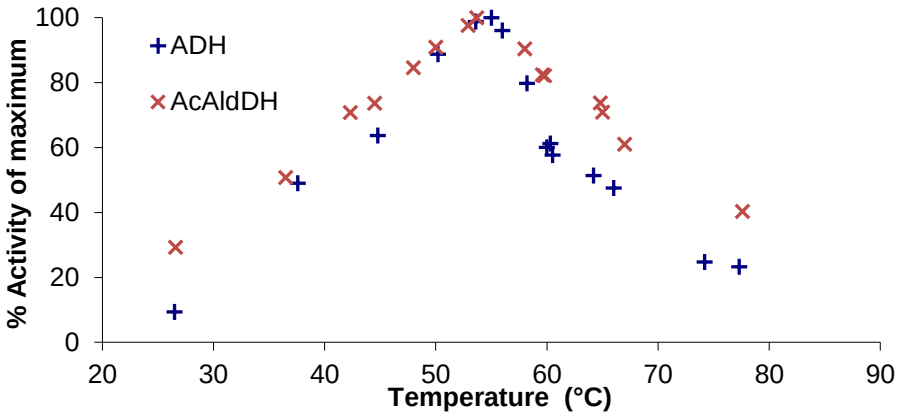


Figure 8-13: Plot of % activity of maximum against temperature of assay (°C) for the two enzymatic activities of Fusion 1.

The optimum temperatures for the two domains of the Fusion 1 protein were both approximately 55°C.

Thermostability assays were carried out on the purified Fusion 1 protein as described in Section 2.13.6. The protein was divided into 50 µl aliquots for incubation at the desired temperatures.

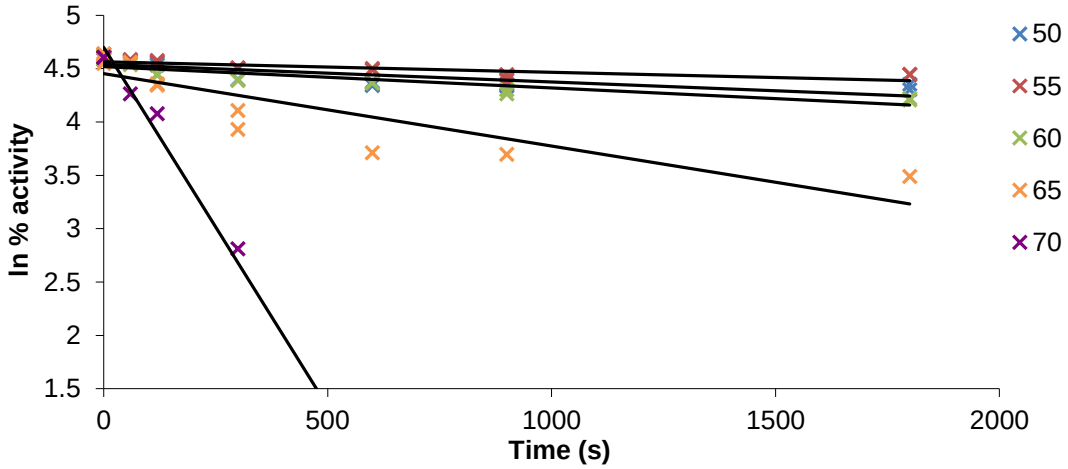


Figure 8-14: Plot of ln % activity against time (s) at a defined temperature (°C) for the AcAldDH activity of Fusion 1.

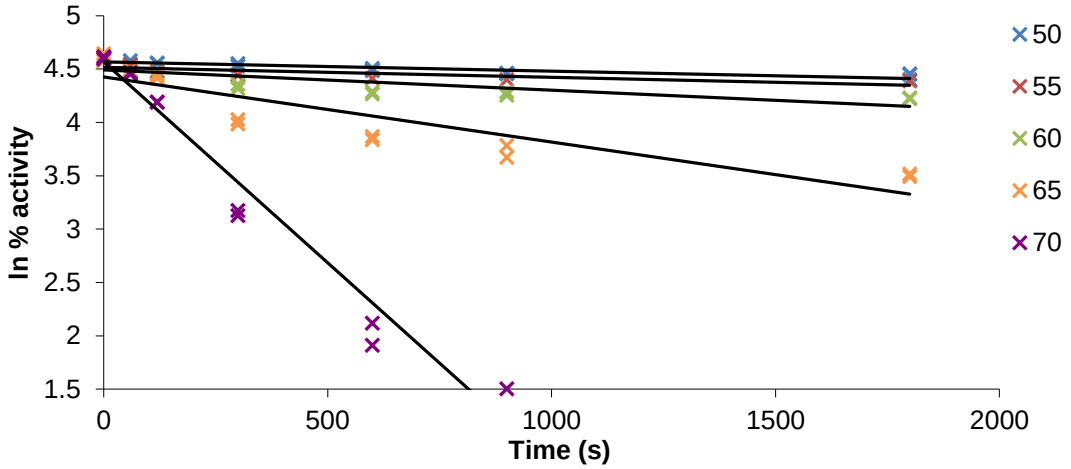


Figure 8-15: Plot of ln % activity against time (s) at a defined temperature (°C) for the ADH activity of Fusion 1.

The two domains of the Fusion 1 protein retained 68% of their original activity following 30 min incubation at 60°C. The half-life of the two domains at 65°C was approximately 5 min.

8.3.1.4 Fusion 1: estimation of M_r by gel filtration

Gel filtration was carried out on the purified Fusion 1 protein in 50 mM Tris-HCl pH 8.0, 150 mM NaCl. Assays of activity were used to monitor the presence of the Fusion 1 protein. The retention time of the peaks of activity were compared to standard proteins to provide an estimate of the M_r.

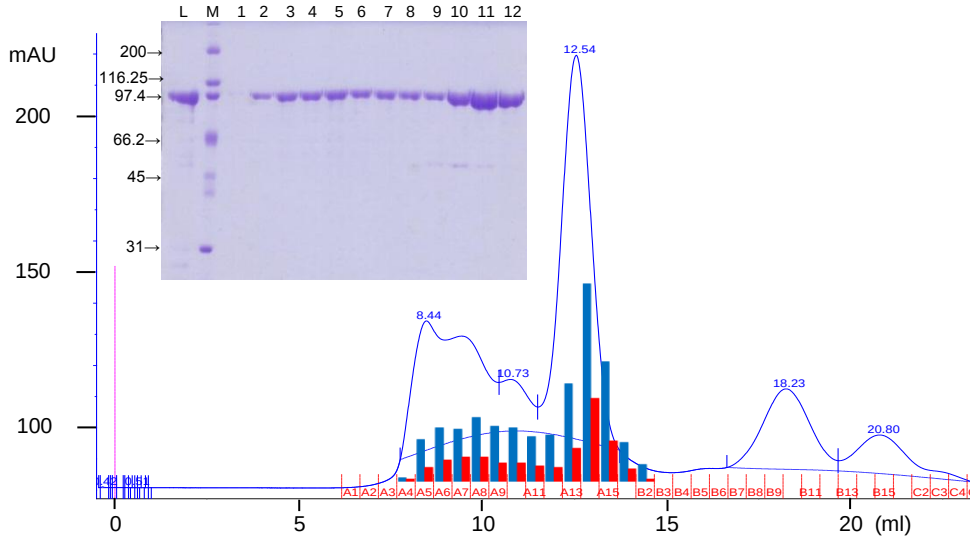


Figure 8-16: Chromatograph of a purified Fusion 1 gel filtration run. Blue line = A_{280nm} (mAU). Activity measurements (abs/min) for both ADH and aldDH are overlaid onto corresponding fractions (blue = ADH, red = aldDH). An SDS-PAGE gel of the peak fractions is also shown. L = load sample, M = markers (M_r/1000) 1-12 = fraction numbers A4-A15 (approximate elution volume 8 ml – 13.5 ml, 0.5 ml per fraction).

Protein peak identity		$V_{elution}$ (ml)	$M_{r\ Observed}$	$M_{r\ Subunit}$	$M_{r\ Observed}$/ $M_{r\ Subunit}$
Fusion 1	peak 1	9.9	610,500	97,705	6.25
Fusion 1	peak 2	12.54	194,000	97,705	1.99

Table 8-5: Estimation of M_r of Fusion 1 from the gel filtration results.

The Fusion 1 protein appeared to be poly-dispersed, eluting over a broad range of sizes. The two major peaks of activity correspond to associations of 6 and 2 polypeptides.

8.3.2 Fusion 1 crystallisation trials

His-purified Fusion 1 was selected for crystallisation trials in an effort to determine how the two domains may interact, and to suggest possible points of modification to improve the activity of the protein *in vivo.*

The protein at 14 mg/ml (dialysed into 50 mM Tris-HCl pH 8.0, 150 mM NaCl and 0.1 mM zinc acetate) was screened at two different ratios of protein to well solution (1:1 and 2:1 protein: well solution, 300 nl drop size) with Structural screen I & II, JCSG-plus and PGA screens (Molecular Dimensions, UK). No hits were obtained for this protein under these conditions.

8.3.3 Fusions 2-4

8.3.3.1 Fusions 2-4 gene cloning

Restriction digests of the final expression constructs of the three Fusion proteins with the domains in the same orientation as the native enzyme are shown in Figure 8-17. The pET28a-*fusion 2* construct was used for expression of the His-tagged Fusion 2 protein in *E. coli*. Later experiments expressing the protein in *Geobacillus* strains used the un-tagged pUCG18-pLDH-*fusion 2* construct. Due to observed susceptibility of the Fusion 2 protein to proteolysis, Fusions 3 and 4 were created. Fusions 3 and 4 were not selected for expression in *Geobacillus* strains.

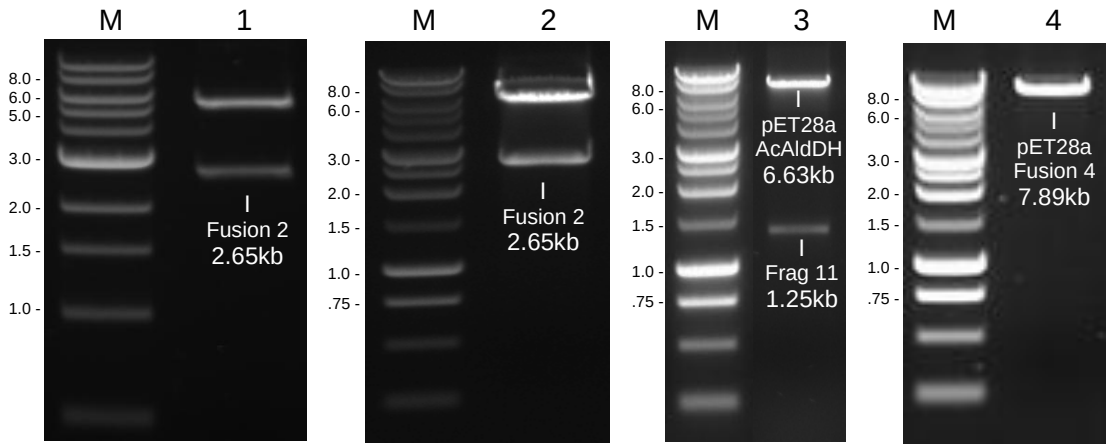


Figure 8-17: Agarose gel electrophoresis of restriction-digested constructs. The predicted sizes of the various DNA fragments are shown over the respective lane for each digest. M = DNA markers sizes given in kb. 1 = pET28a-*fusion 2* (digested with NheI/XhoI), 2 = pUCG18-pLDH-*fusion 2* (digested with XbaI/SacI), 3 = pET28a-*fusion 3* (digested with HindIII/XhoI) & 4 = pET28a-*fusion 4* HindIII site removed (digested with HindIII/XhoI).

8.3.3.2 Fusion 2: protein expression and characterisation

The purified plasmid construct was transformed into the *E. coli* BL21 (DE3) protein expression strain. Protein expression was assessed following 4.5 h of protein expression conditions. Protein of the predicted M_r was observed in the soluble cell extract under these conditions, and was purified using metal affinity chromatography (column charged with Ni^{2+} according to the method outlined in Section 2.9). Analysis of the purification samples by SDS-PAGE and enzyme assays of the protein-containing fractions are shown in Figure 8-18 & Figure 8-19, respectively.

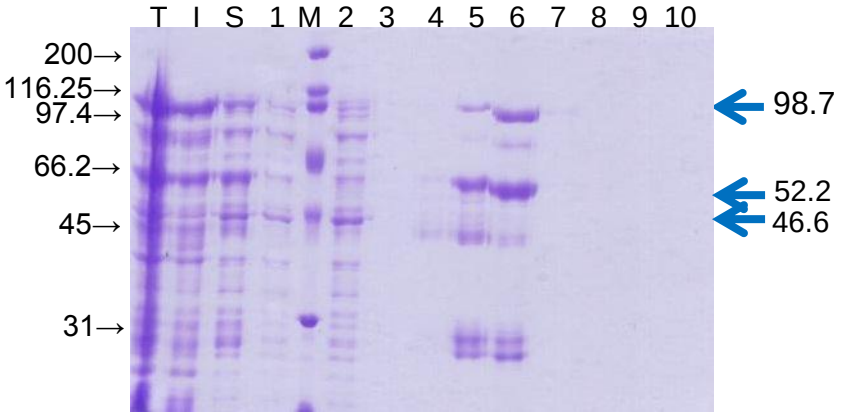


Figure 8-18: SDS-PAGE analysis of Fusion 2 metal affinity chromatography samples. T = total, I = insoluble, S = soluble, 1 = flow through, M = markers (M_r/1000), 2 = 0%, 3 = 2.5%, 4 = 5%, 5 = 10%, 6 = 20%, 7 = 30%, 8 = 40%, 9 = 60%, 10 = 100% (% values = %HIS-ELUTE buffer diluted in HIS-BIND buffer). Predicted His-tagged-Fusion 2 band (M_r = 98,712) is highlighted with a blue arrow. The size of the two protein fragments that would result from proteolysis, assuming cleavage occurred exactly between the two domains, are also highlighted with blue arrows (M_r/1000 is indicated).

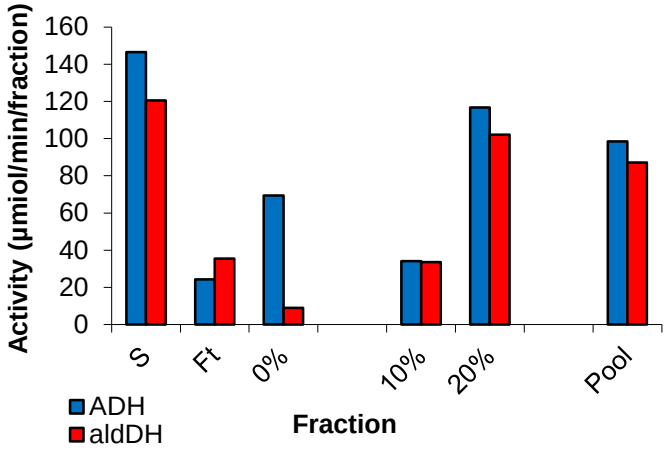


Figure 8-19: Enzyme activity assay results (U sample^{-1}) for key fractions of the Fusion 2 metal affinity purification. S = soluble, Ft = flow through, % values = % wash step & Pool = pooled 10 & 20% sample.

As shown in Figure 8-19, the relative activities of the two enzymes present in the fusion samples varied between purification fractions. SDS-PAGE gel analysis showed a number of other proteins present in the samples, some co-eluting with the purified Fusion 2 protein. It was hypothesised that proteolysis between the two domains resulted in two functional enzymes, the His-tagged AcAldDH (seen to co-elute with Fusion 2) and the un-tagged ADH domain (observed in the 0% wash sample). The 10 & 20% fractions were therefore enriched with AcAldDH activity, whereas the 0% wash sample was enriched with ADH activity.

In an effort to further resolve the hypothesised AcAldDH-containing degradation product, a pooled 10 & 20% fraction from the metal affinity purification was run on a gel filtration column.

8.3.3.3 Fusion 2: purification by gel filtration

Gel filtration was carried out on the purified Fusion 2 protein in 50 mM Tris-HCl pH 8.0, 150 mM NaCl. Assays of activity were used to monitor the presence of the protein of interest. The retention time of the peaks of activity were compared to standard proteins to provide an estimate of the M_r.

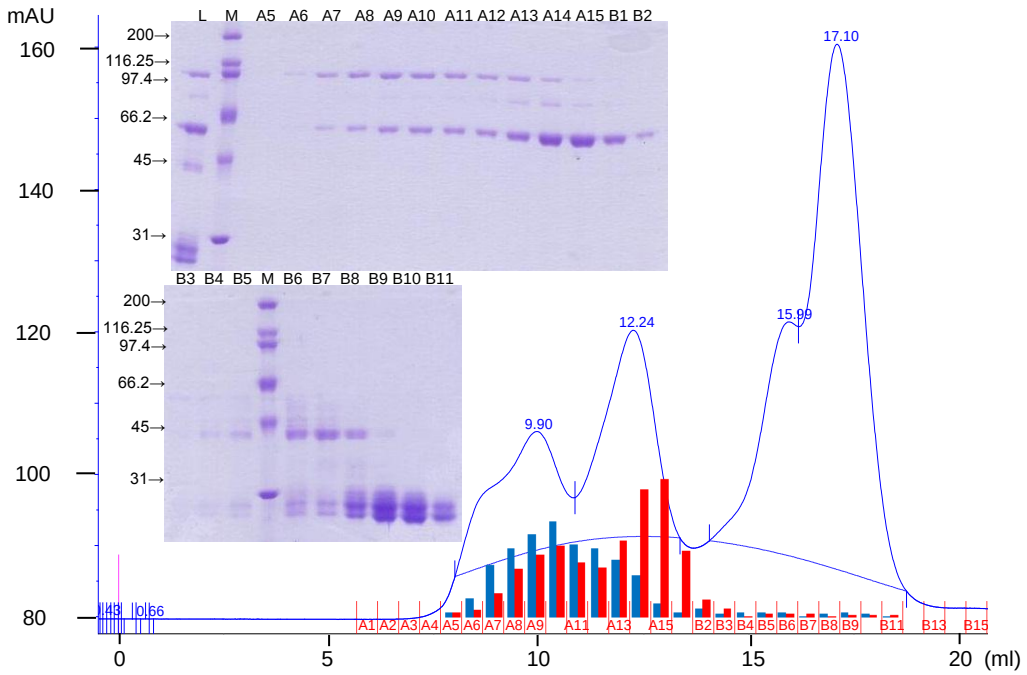


Figure 8-20: Chromatograph of a partially-purified Fusion 2 gel filtration run. Blue line = A_{280nm} (mAU). Activity measurements (abs/min) for both ADH and aldDH are overlaid onto corresponding fractions (blue = ADH, red = aldDH). SDS-PAGE gels of the peak fractions are also shown. L = load sample, M = markers (M_r/1000) & A5-B11 = fraction numbers.

Two enzymatically active protein species were eluted from the gel filtration column. The first peak of enzyme activity contained both ADH and AcAldDH activity corresponding to the Fusion 2 protein. A second peak of AcAldDH activity eluted shortly after; the protein responsible for this peak appeared to be the predicted degradation product of the Fusion. ADH activity independent of AcAldDH was not detected, presumably because the ADH degradation product was eluted in the flow through and 0% wash step of the previously-run metal affinity purification. The AcAldDH degradation product was not fully resolved from the Fusion 2 protein by gel filtration.

SDS-PAGE analysis of the gel filtration fractions (Figure 8-20), gave an estimate of M_r for the dominant protein species in the AcAldDH containing fraction of 54,900. This value for M_r was used for the prediction of AcAldDH subunit assembly.

Protein peak identity		$V_{elution}$ (ml)	$M_{r\ Observed}$	$M_{r\ Subunit}$	$M_{r\ Observed}$/ $M_{r\ Subunit}$
Fusion 2	peak 1	10.5	470,000	98,712	4.8
AcAldDH?	peak 2	12.5	198,000	54,900	3.6

Table 8-6: Peak analysis of Fusion 2 purification, M_r estimates for peak of protein activity are shown.

8.3.3.4 NADH→NAD: CoA released ratio assays

Assays to determine whether the Fusion 2 protein was capable of catalysing the conversion of acetyl-CoA to ethanol were carried out as described in Section 2.13.4. As observed for the Fusion 1 protein, an average of three assays gave a ratio of 1.2:1 NADH converted to CoA-SH released. Neither of the Fusion proteins appeared to be able to catalyse conversion of acetyl-CoA to ethanol *in vitro.*

Due to the observed instability of the Fusion 2 protein and the difficulty in fully resolving the Fusion protein from high levels of degradation products, this Fusion was not kinetically characterised.

8.3.3.5 Fusions 3 & 4

The susceptibility of the Fusion 2 protein to proteolysis was thought to be due to the length of the amino acid sequence between the two protein domains. The Fusion 2 protein contained a longer "linker" region than the native ADHE protein, as illustrated in Figure 8-21. Two further Fusion proteins were created with shorter "linker" sequences between the two domains of the protein (as described in Sections 8.2.2 and 8.3.3.1).

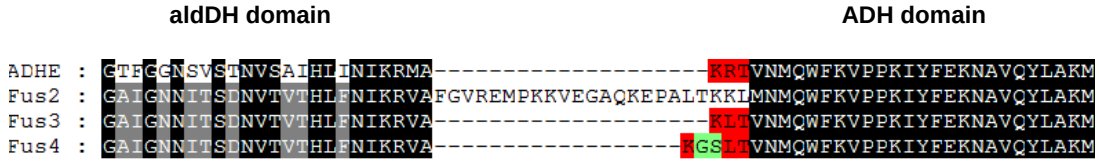


Figure 8-21: Protein sequence alignment between the native ADHE protein and the 3 native orientation Fusion proteins (Fus = Fusion). Identical residues highlighted in black and grey, "linker" sequence is indicated in red, and additional GS insertion is indicated in green.

As this work was started prior to the resolution of the crystal structure of the AcAldDH protein, the protein structure prediction programme SWISS-MODEL (ExPASy (Kiefer et al. 2009)) was used to create a structural homology model of the AcAldDH domain. This was based on the aldDH domain of an ADHE from *Vibrio parahaemolyticus* (41% identity, 61% similarity, 5% gaps; PDB: 3MY7). This model was used to guide the

shortening of the linker region, avoiding the loss of key structural features possibly required for protein activity.

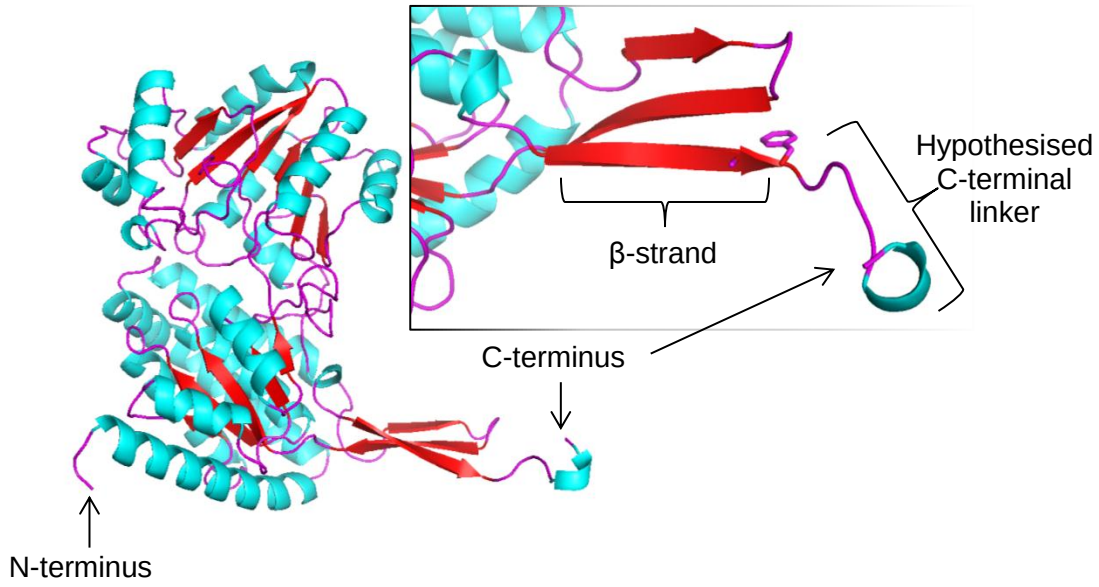


Figure 8-22: AcAldDH predicted protein structure from SWISS-MODEL. Red arrows = β strands, Blue spirals = α helices. The N and C termini are indicated as well as the hypothesised linker region.

Sequence alignment showed the C-terminal NIKRxA motif was conserved between the AcAldDH and the aldDH domain of ADHE. The model predicted this motif to be part of a C-terminal β strand in the AcAldDH protein. This strand was hypothesised to be the end of the AcAldDH coding region. The subsequently-resolved AcAldDH crystal structure (Chapter 7) confirms this residue as the end of the structured region of the protein. The linker was truncated to the Alanine residue (467 in the His-tagged Fusion 2 protein) at the C-terminal end of the β-strand (the terminal residue of the AcAldDH domain in Figure 8-21) to give Fusion 3. Fusion 4 incorporated two additional amino acids (G and S) into the linker to increase its length slightly.

8.3.3.6 Fusion 3 and 4: protein expression

The purified plasmid constructs were transformed into the *E. coli* BL21 (DE3) protein expression strain. Proteins of the predicted M_r were observed in the soluble cell extracts following 4.5 h of protein expression conditions, and were purified using metal affinity chromatography (according to the method outlined in Section 2.9). Analysis of the purification samples by SDS-PAGE and enzyme assays of the protein-containing fractions are shown in Figure 8-23 & Figure 8-24.

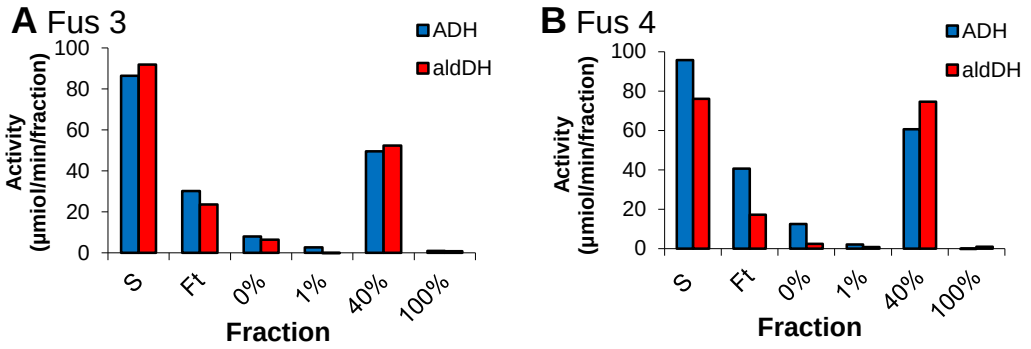


Figure 8-23: Enzyme activity assays (U sample^{-1}) for fractions of the Fusions 3 & 4 metal affinity purifications (A = Fusion 3 B = Fusion 4). S = soluble, Ft = flow through & % values = % wash step.

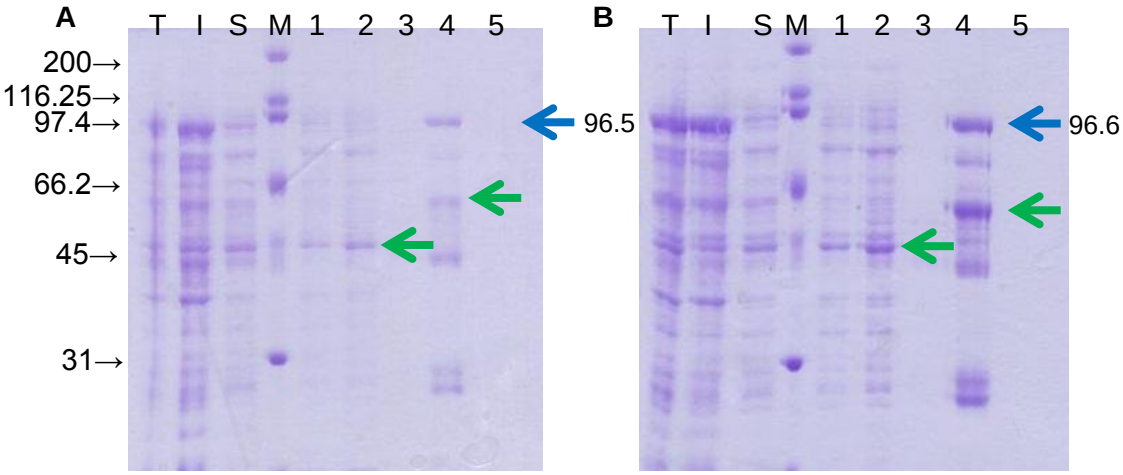


Figure 8-24: SDS-PAGE analysis of Fusions 3 & 4 metal affinity chromatography samples (A = Fusion 3 (M_r = 96,455); B = Fusion 4 (M_r = 96,599)). T = total, I = insoluble, S = soluble, M = markers (M_r/1000), 1 = flow through, 2 = 0%, 3 = 1%, 4 = 40% & 5 = 100% (% values = %HIS-ELUTE buffer diluted in HIS-BIND buffer). Predicted His-tagged-Fusion bands are highlighted with a blue arrow (M_r/1000). Probable proteolysis products are highlighted with green arrows.

Results of assays and SDS-PAGE analysis showed the susceptibility to proteolysis between the domains remained despite a reduction in the sequence length of the linker. No further attempts were made to increase resistance to proteolysis.

8.3.4 Fusions 1 & 2: *in vivo* evaluation

Assays of Fusions 1 & 2 *in vitro* showed that the proteins were unable to catalyse the conversion of acetyl-CoA to ethanol under assay conditions. Experiments to evaluate the Fusion proteins *in vivo* were undertaken to see if expression under fermentative (anaerobic) conditions would allow conversion to ethanol to occur.

The TM400 strain (ADHE-knockout) is unable to grow under fermentative conditions. This strain was transformed with pUCG18-pLDH constructs coding for the native

ADHE, Fusions 1 & 2, as well as the Fragment 11 and AcAldDH proteins; screening used the *kan* resistance marker gene in this vector.

8.3.4.1 Tube fermentation experiments

Tube fermentations were carried out for the various TM400-pUCG18-pLDH strains on USM media containing 100 mM glucose and 0.5% yeast extract, as described in Section 2.5.1. Kanamycin (12 µg/ml) was present for all strains transformed with the pUCG18-pLDH constructs. This initial experiment evaluated the ability of the Fusion proteins and their component domains to restore to TM400 the ability to grow and produce ethanol, under fermentative conditions.

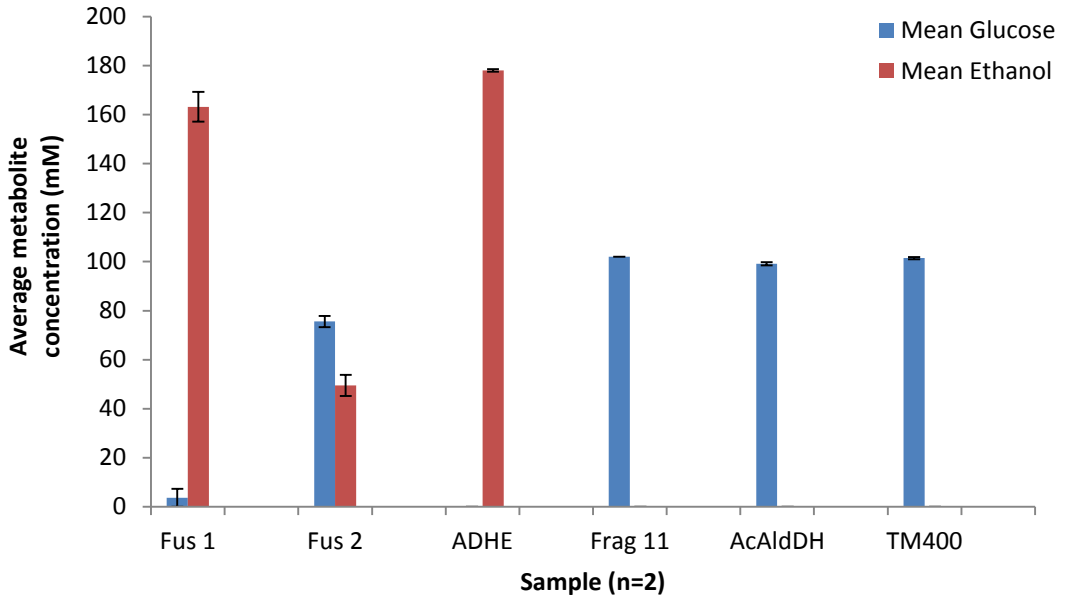


Figure 8-25: Metabolite analysis of 48 h tube fermentation experiment. Each sample is the average of 2 fermentation tubes (error bars = standard deviation from mean). Data for ethanol and glucose are shown.

A second tube fermentation experiment sought to confirm the observations of the first experiment, with an increased number of samples but also comparing the results to the TM242 strain grown under identical conditions.

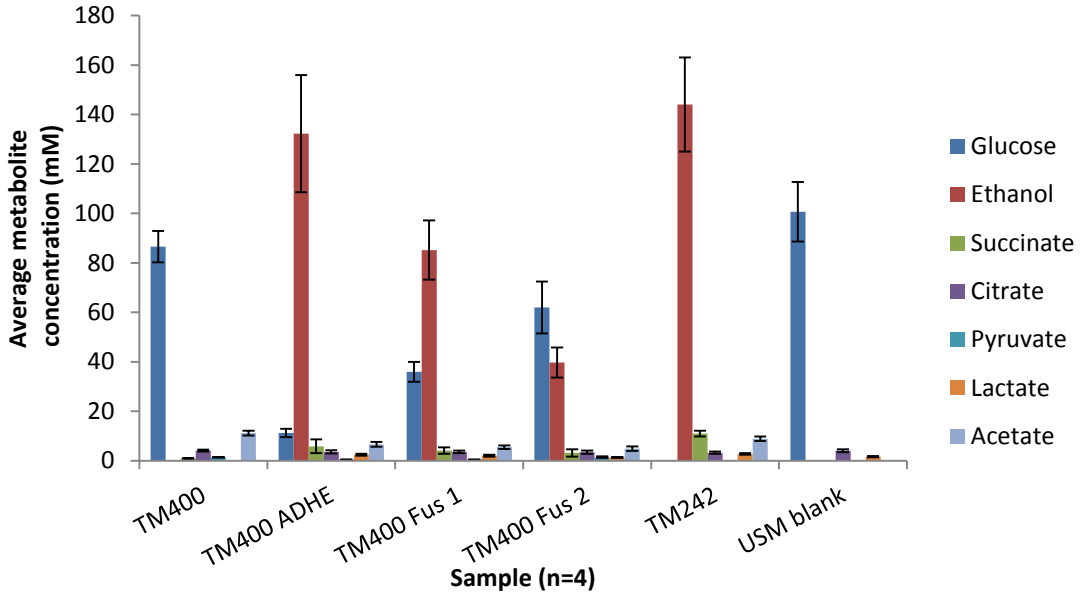


Figure 8-26: Metabolite analysis of 48 h tube fermentation experiment. Each sample is the average of 4 fermentation tubes (error bars = standard deviation from mean).

Expression of ADHE using the pUCG18-pLDH vector confirmed the phenotype of ethanol production could be rescued using proteins expressed in this manner (the strain is now comparable to TM242). The AcAldDH and Fragment 11 proteins in isolation (produced in previous chapters) were shown to be incapable of rescuing the ethanol production phenotype. Fusions 1 and 2 were both able to rescue the phenotype and to catalyse the production of ethanol *in vivo*. Yields of ethanol produced for the Fusion proteins did not reach the level of TM242 or TM400-ADHE. The highest yield for the Fusions was observed for the Fusion 1 protein.

8.3.4.2 NADH→NAD: CoA released ratio assays

The ability of the Fusion proteins to rescue the phenotype of ethanol production in the tube fermentation experiments was in contrast to the result of *in vitro* ratio assays carried out on the recombinant proteins. Further assays were therefore undertaken to determine if the Fusion 1 protein expressed using the *Geobacillus* system (no His-tag), was different from that produced in the *E. coli* strains used in Section 8.3.1.2. Cell extracts were produced from TM400 strains transformed with the Fusion 1 and ADHE constructs and grown under tube fermentation conditions for 24 h (5 tubes pooled to collect sufficient biomass). Extracts were assayed using the method described in Section 2.13.4.

Assays of the cell extract of the Fusion 1 sample gave a ratio of 0.95:1 NADH converted to CoA-SH released (average of two assays), in contrast to the ADHE

sample which gave a ratio of 1.7:1 NADH converted to CoA-SH released (average of two assays). Thus the native ADHE protein appeared able to catalyse ethanol production *in vitro* whereas the Fusion 1 protein could not.

8.3.4.3 Fusion 1 and ADHE: full fermentation experiment

To further evaluate the ability of the Fusion 1 protein to catalyse ethanol production *in vivo,* fermentation runs on 2 and 4% glucose were carried out alongside TM400-ADHE (positive control). The fermentations were carried out under the conditions defined in Table 8-7.

Fermentation parameter	**TM400 ADHE #1**	**TM400 ADHE #2**	**TM400 Fus 1 #1**	**TM400 Fus 1 #2**
% Glucose	2	4	2	4
Media	50 mM USM, X% Glucose (shown above) 2% Yeast Extract 2x AA mix 1			
Working volume (L)	1			
Inoculum (%v/v)	100 ml (10%) of an OD_{600nm} 8.7 seed culture		100 ml (10%) of an OD_{600nm} 8.0 seed culture	
Temperature (°C)	60			
pH	6.7			
Agitation (RPM) pre "anaerobic switch"	600			
Aeration (L/min) pre "anaerobic switch"	1			
Anaerobic switch (OD_{600nm})	5	5.1	5	4
Agitation (RPM) post "anaerobic switch"	300			
Aeration (L/min) post "anaerobic switch"	0.2			

Table 8-7: Experimental methodology of TM400-ADHE vs TM400-Fusion 1 (Fus 1) fermentations.

The following figures summarise the fermentation runs carried out for the two constructs. Wherever possible, conditions used for the two strains were identical.

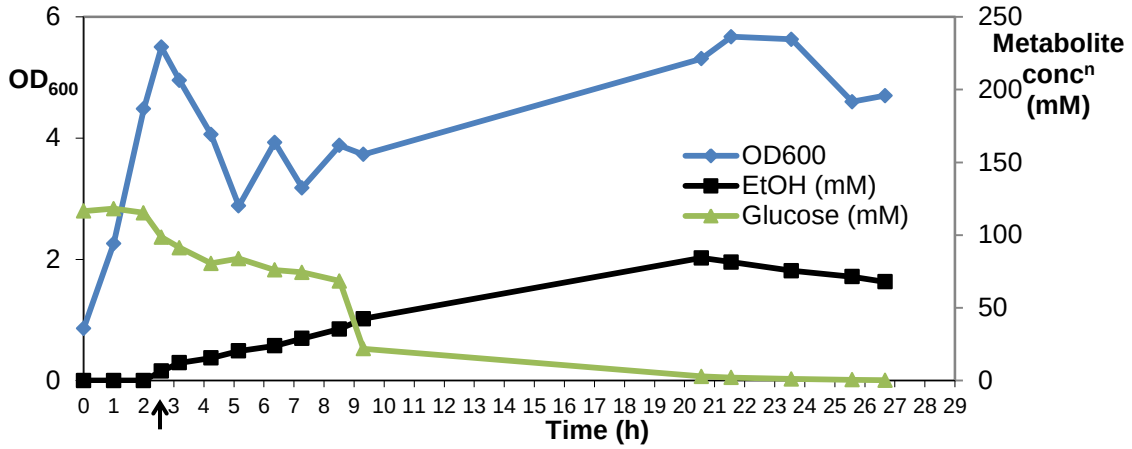


Figure 8-27: Fermentation summary for TM400-ADHE with 2% glucose in the starting media. Black arrow indicates anaerobic switch.

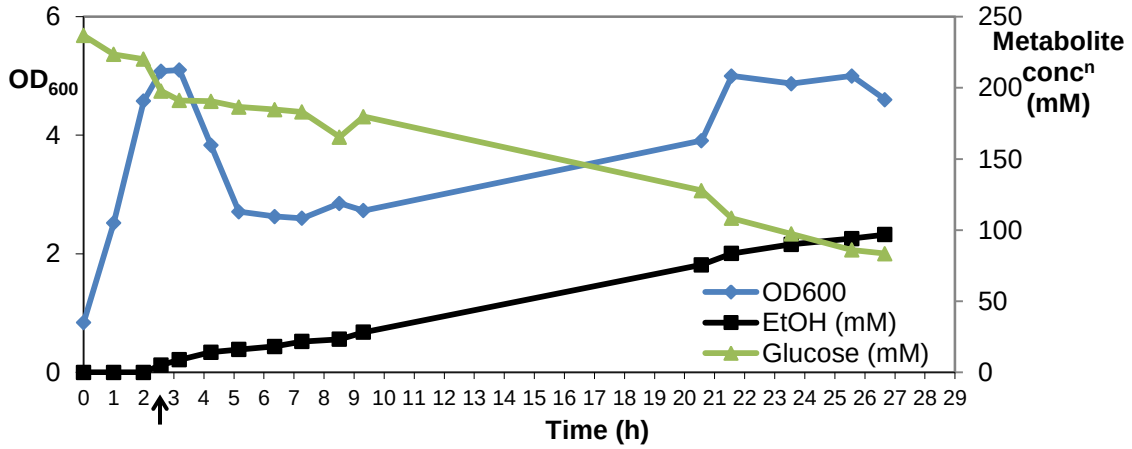


Figure 8-28: Fermentation summary for TM400-ADHE with 4% glucose in the starting media. Black arrow indicates anaerobic switch.

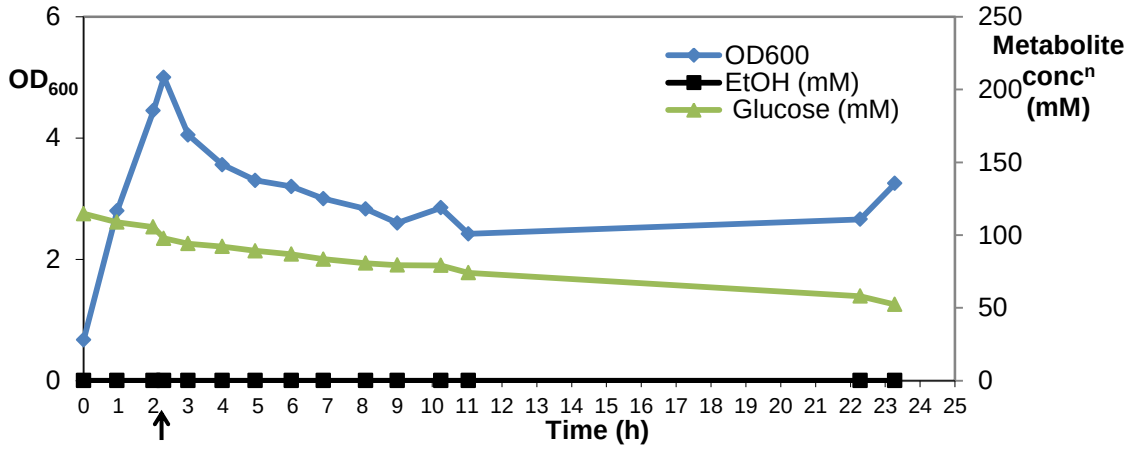


Figure 8-29: Fermentation summary for TM400-Fus1 with 2% glucose in the starting media. Black arrow indicates anaerobic switch.

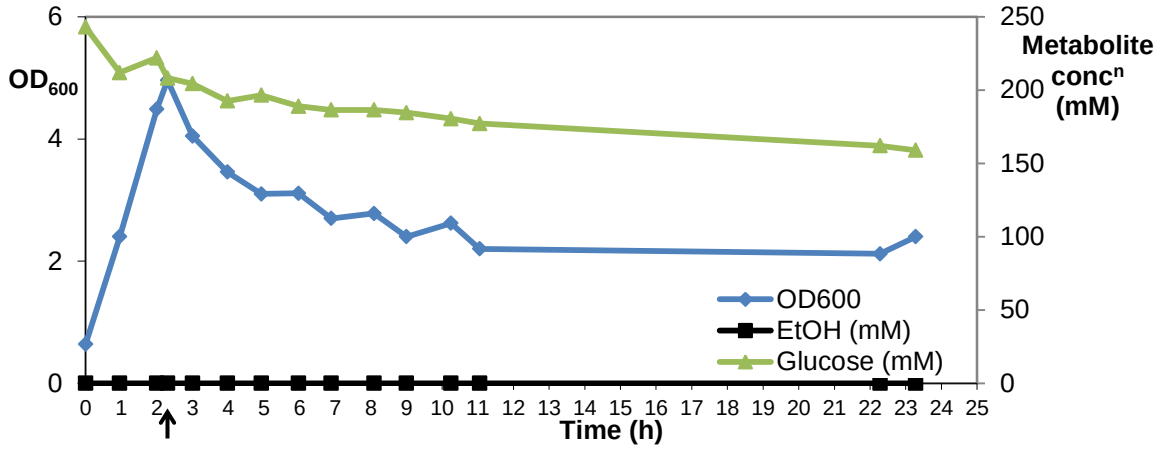


Figure 8-30: Fermentation summary for TM400-Fus1 with 4% glucose in the starting media. Black arrow indicates anaerobic switch.

The TM400-ADHE strain was able to produce ethanol under fermentative conditions in a similar manner to TM242 on both 2% and 4% glucose. The observed drop in OD_{600nm} post anaerobic switch, followed by an increase as metabolism progresses, is typical of TM242 fermentations (TMO Renewables 2012 personal communication). The TM400-Fusion 1 strain was unable to metabolise glucose under fermentative conditions. The observed drop in OD_{600nm} post anaerobic switch was not followed by any recovery for this strain.

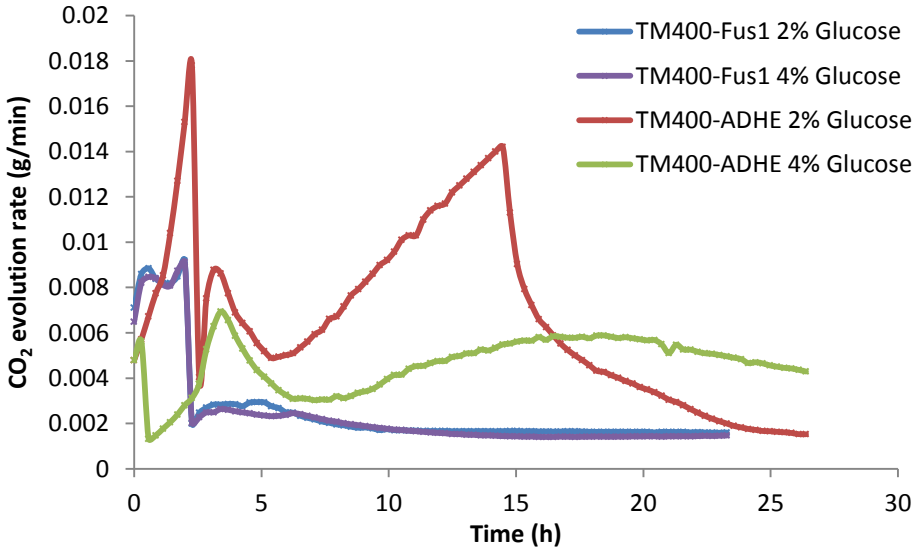


Figure 8-31: CO_2 evolution rates measured using in-line mass spectrometry for the TM400-Fus1 and TM400-ADHE fermentations. Anaerobic switch made at 2.3 h for TM400-Fus1 strains and 2.6 h for TM400-ADHE strains.

CO_2 evolution rates are commonly used as a measure of metabolic activity. CO_2 is primarily produced by the PDH complex, so increases in CO_2 evolution after the anaerobic switch indicate an increased rate of fermentative metabolism. The initial

decrease observed for all strains in Figure 8-31 is due to both the lowering of the aeration rate post anaerobic switch and the switch from aerobic to fermentative metabolism. The increase post switch observed for TM400-ADHE is similar to the observed pattern in TM242, where metabolic flux to ethanol is increasing. The TM400-Fusion 1 strain was unable to recover post anaerobic switch. Initiation of sporulation is a likely explanation for the lack of ethanol production in this strain and, if correct, was probably triggered by the switching of conditions from aerobic to "anaerobic".

8.4 Discussion

Artificial ADHE proteins were produced in this section of the project by creating fusions of the previously-characterised Fragment 11 and AcAldDH proteins. Fusions of the domains in two different orientations were produced, and in both the ADH and AcAldDH were active. However, the fusion of Fragment 11 with EutE protein resulted in an insoluble protein.

The Fusion 1 protein (N-ADH-AcAldDH-C; i.e. opposite orientation to the native ADHE protein) appeared to be more stable in terms of proteolytic degradation than the fusions created in the native orientation (N-AcAldDH-ADH-C). This Fusion 1 was shown to be thermostable at the growth temperatures of TM242 and both domains of the protein were shown to be optimally active at 55°C. Fusion of the more thermostable AcAldDH (half-life greater than 30 min at 60°C) to the less thermostable Fragment 11 (half-life of 10 min at 60°C), appeared to improve the thermostability of the Fragment 11 protein. The two domains of the Fusion 1 protein retained 68% of their original activity following 30 min incubation at 60°C. The mobile loops shown in the Fragment 11 crystal structure may be stabilised by the presence of the AcAldDH protein, thus explaining the improved thermostability observed.

Co-expression of the two activities in the Fusion 1 protein caused changes in the kinetic parameters observed, as shown in Table 8-8. As all the proteins were purified to virtual homogeneity the parameters can be directly compared. As discussed in Chapter 3 values of k_{cat}/K_m should be treated with caution as the rates observed will depend on the physiological concentrations of substrates.

Activity	Substrate	$V_{max\ (adjusted)}$ (U mg^{-1})	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1} \bullet s^{-1}$)
Fusion 1					
ADH	Acetaldehyde	145.9	151.0	237.5	1.6
	NADH	110.6	0.131	180.1	1375.1
AcAldDH	Acetyl-CoA	30.2	0.032	49.2	1538.7
	NADH	28.2	0.057	45.9	806.4
Fragment 11					
ADH	Acetaldehyde	430.8	121.0	349.0	2.9
	NADH	427.2	0.062	346.1	5582.0
AcAldDH					
AcAldDH	Acetyl-CoA	37.5	0.024	32.6	1357.6
	NADH	34.5	0.042	30.0	713.7

Table 8-8: Summary of the catalytic properties of the Fusion 1 protein and its constituent domains expressed independently.

In terms of k_{cat}, the AcAldDH protein appeared more catalytically active in the Fusion than when expressed independently. The opposite was true for the Fragment 11 domain where the k_{cat} of the ADH activity was significantly lower in the Fusion than in isolation. The K_m values observed for the various substrates of the domains were similar for both the Fusion and the isolated proteins. The significant exception to this is the K_m value for Fragment 11 with respect to NADH. The AcAldDH at the C-terminus of the protein (Fusion 1) appears to cause an increase in the K_m observed for the ADH in the Fusion, compared to when expressed in isolation. As shown in Chapter 5, the C-terminus of the Fragment 11 protein is positioned such that the AcAldDH domain could interact with the predicted active site of Fragment 11. This interaction may affect the activity of the proteins involved. Competition for NADH between the two domains may also explain why the K_m for this substrate increases so significantly.

Proteolytic degradation was observed between the two domains of the native orientation Fusion proteins (2, 3 and 4). Although different peptide linkers were tested for this orientation, the resulting fusions all appeared to remain susceptible to degradation. Considering the now resolved Fragment 11 crystal structure, the unstructured region at the N-terminus of the protein is a possible site of proteolysis between the two domains in the native orientation Fusions. As the N-terminal His-tag in the Fusion 1 protein was not cleaved in this region, and the native ADHE protein was not rapidly proteolysed this is considered less likely. Future investigations could focus on the effect of truncating the N-terminal region of the Fragment 11 protein, to

investigate this region in terms of protease sensitivity. Currently, the susceptibility of Fusions 2, 3 and 4 to degradation remains unresolved. Due to the issues resolving the Fusion proteins from the AcAldDH fragment produced by inter-domain proteolysis, kinetic parameters were not determined for these fusion proteins.

Evaluation of the ability of the Fusion proteins to catalyse the conversion of acetyl-CoA to ethanol was tested. Neither of the orientations of the two domains appeared to be able to carry out the two reactions consecutively. Factors such as the product inhibition observed for the AcAldDH protein in Chapter 6, and the high K_m for acetaldehyde observed for the ADH domain, may be limiting flux to ethanol.

Evaluation of the Fusion proteins *in vivo* using the tube fermentation method, showed the proteins to be capable of restoring the phenotype of ethanol production to an ADHE knock out strain TM400. The Fusion 1 protein initially appeared capable of restoring comparable yields of ethanol to the ADHE protein expressed using the same construct. The less stable Fusion 2 protein was shown to produce significantly lower yields of ethanol than the ADHE protein and the Fusion 1 protein. The Fusion 1 protein was selected for further evaluation using standard fermentation methods.

The fermentation experiments gave contrasting results to the tube fermentations. The Fusion 1 containing strains were unable to metabolise glucose to ethanol under fermentative conditions. The strains expressing the native ADHE protein using the same construct were able to successfully metabolise glucose to ethanol under identical conditions. It would therefore appear that limitations in the activity of the Fusion 1 protein prevented successful fermentation by the TM400-Fusion 1 strain. A possible explanation for the differences observed between these experiments was the nature of the anaerobic switch. Under tube fermentation conditions, the switch to fermentative metabolism is controlled by the level of oxygen decreasing due to aerobic growth of the organism. The oxygen in the tube is gradually used up which induces the switch in metabolic pathways. This gradual switch may be considered much less of a “shock” to the organism compared to under fermenter controlled conditions. Here the switch is forced through the decreased aeration and stirring of the culture. The switch is a sudden change, which may be capable of triggering the stress response of sporulation in the organism. Unfortunately, no sporulation deficient ADHE knock out strain was available to test this hypothesis. Tube fermentations are used as an established high throughput screening method by TMO Renewables to characterise ethanol production of various strains. However, it is not unusual for strains to behave differently between

the initial screen and full fermentation evaluation (TMO Renewables 2012 personal communication).

The tube fermentation results demonstrated that under the correct metabolic conditions, the Fusion proteins were capable of catalysing ethanol production in the ADHE knock out strain TM400. The ability of the TM400 strain to produce ethanol was not restored through expression of the Fragment 11 or AcAldDH proteins in isolation. It is therefore tempting to suggest this as an indication that the ADHE protein is required for both enzymatic activities, rather than being supplemented by other proteins present in the cell extract (for example the other annotated ADH coding genes). Due to the previously described limitations of the protein domains used, it would be unwise to draw such a conclusion. Whether the key activities required for the conversion of acetyl-CoA to ethanol are indeed solely coded for by the ADHE protein remains un-elucidated.

A UK patent application (GB1114701.4) was filed in conjunction with TMO Renewables detailing the potential use of the Fusion proteins created here, in terms of their use in modulating ethanol production.

To summarise, Fusion proteins comprising the AcAldDH and Fragment 11 proteins have been produced in a soluble form, presenting both aldDH and ADH activities. Co-expression of the two domains as a single protein does not appear to resolve key limitations of the individual component domains. Preliminary tube fermentation experiments showed both orientations of the Fusion proteins to be capable of restoring the ability of the TM400 strain to produce ethanol. Full fermentation evaluation did not demonstrate the same result. Differences in the initiation of fermentative metabolism between these two experiments are a possible explanation for the differences in the results observed.

9 INVESTIGATION INTO THE EXPRESSION OF KEY METABOLIC ENZYMES THROUGH A FERMENTATION RUN

9.1 Introduction

TMO Renewables have engineered the organism *G. thermoglucosidasius* to produce the TM242 strain optimised for ethanol production. This strain has been shown to fermentatively metabolise glucose to ethanol, with yields of ethanol approaching 90% of the theoretical maximum (0.51 g ethanol /g glucose) (Cripps et al. 2009). The manipulated metabolic pathway in fermentative glucose metabolism of TM242 is shown in Figure 9-1. Efficient metabolism of the carbon source is controlled by the metabolic enzymes present in the organism. Directing the metabolic flux to the desired end products is influenced by a range of factors, including the relative expression levels of enzymes, enzyme kinetic properties and metabolite concentrations.

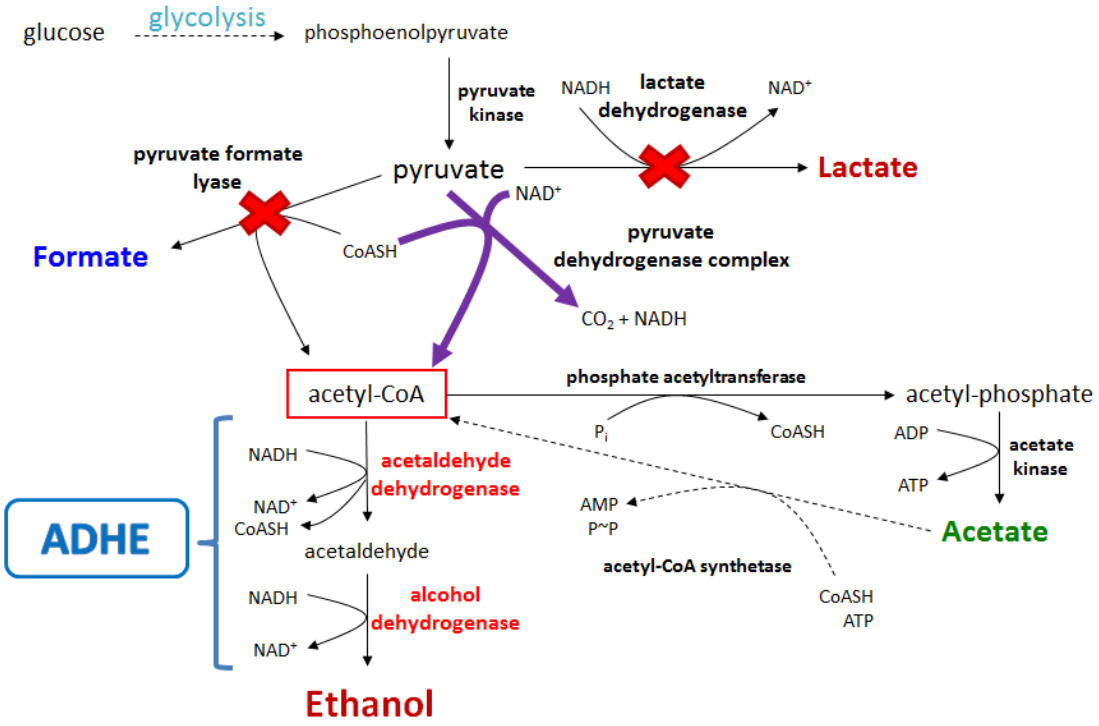


Figure 9-1: An overview of fermentative metabolism in *Geobacillus thermoglucosidasius* TM242. Red crosses indicate gene knock-outs and purple arrows indicate up-regulation through promoter switching. The dashed arrow for acetyl-CoA synthetase indicates a hypothesised activity yet to be confirmed.

Under fermentative conditions two relevant competing enzyme activities exist for the consumption of acetyl-CoA in TM242. These are the aldDH activity associated with ADHE and the phosphate acetyltransferase (PAT). As mentioned previously, preliminary data suggest that PAT may be a more catalytically active enzyme than the

aldDH domain of ADHE (Hills, C 2011 unpublished work). If this is the case *in vivo* these two enzyme activities must be under some form of regulation, to allow ethanol production to proceed to the yields observed.

This chapter describes work carried out to investigate the expression and kinetic properties of the enzymes controlling ethanol production under standardised conditions. This work includes the determination of enzyme expression patterns through a fermentation run.

The work reported in this part of the project was the result of a combined effort of the author, Christopher Hills and TMO Renewables. Christopher is a fellow PhD student working in conjunction with TMO Renewables with a focus on the competing branch of metabolism to ADHE. This is made up of the phosphate acetyl transferase (PAT), acetate kinase (AK) and the hypothesised acetyl-CoA synthetase (ACS) enzymes. Data on these three enzymes were therefore determined by him and are included in this chapter with his permission.

TMO Renewables are continuing to modify the TM242 strain to optimise various aspects of metabolism. The TM444 strain was the current production strain at this stage of the project. This strain is identical to TM242 except for one additional modification removing the ability of the organism to sporulate. This further optimised strain was used for the fermentation run carried out during these experiments. The activities associated with the PAT, AK, ADHE and also the acetyl-CoA producing pyruvate dehydrogenase complex (PDH), were all monitored during assays of fermentation samples.

9.2 Materials and methods

9.2.1 Enzyme assays

Unless otherwise stated, enzyme assays were carried out in 50 mM HEPES buffer pH 6.5, 0.1 mM zinc acetate, and with a maintained ionic strength of 0.1 M using NaCl. This buffer was selected as it was important to assay aldDH activity in phosphate-free conditions, preventing PAT interference during the assays. All assays monitored enzyme activities at 60°C using a Varian Cary 50 Bio UV/visible light spectrophotometer with a Peltier temperature controller.

AldDH assays used the DTNB-based assay and ADH assays were carried out using the NADH assay as described in Section 2.13. For the samples from the fermentation

run, the aldDH assays contained 0.35 mM acetyl-CoA and 0.18 mM NADH. ADH assays contained 40 mM acetaldehyde and 0.18 mM NADH.

9.2.1.1 Pyruvate dehydrogenase complex (PDH) activity assays

The assay used for PDH activity was based on assays developed by Dr Tracey Goult (University of Bath) in a postdoctoral research project in conjunction with TMO Renewables. The pyruvate-dependent reduction of NAD^+ in the presence of the TPP cofactor and CoA-SH was measured at 340 nm.

Chemical	Stock conc^n (mM)	Volume added (µl)	Assay conc^n (mM)
KPO_4 (pH 7.0)	53	950	~50
NAD^+	2.9		2.8
Thiamine pyrophosphate (TPP)	0.2		0.2
$MgCl_2$	1.1		1.0
Coenzyme A	6.5	20	0.13
DL-cysteine	130		2.6
Sodium pyruvate	100	20	2
Enzyme / cell extract		10	

Table 9-1: PDH assay conditions.

- Buffer pre-incubated at 60°C
- Substrates stored on ice
- Reaction started by addition of enzyme

9.2.1.2 Phosphate acetyltransferase (PAT) activity assays

The assay used for PAT activity monitored the phosphate-dependent cleavage of DTNB to NTB^{2-} by CoA-SH (Ellman 1959) at 412 nm (ε_{412} = 13,600 $M^{-1}.cm^{-1}$).

Chemical	Stock conc^n (mM)	Volume added (µl)	Assay conc^n (mM)
HEPES buffer pH 6.5, 0.1 mM zinc acetate	50	920	~50
Sodium dihydrogen orthophosphate (dihydrate)	1000	50	50
DTNB	10	10	0.10
Acetyl-CoA	11.51	10	0.12
Enzyme / cell extract		10	

Table 9-2: PAT standard assay conditions.

- Buffer pre-incubated at 60°C
- Substrates stored on ice
- DTNB dissolved in 1 M Tris-HCl pH 8.0
- Reaction started by addition of enzyme
- Ionic strength maintained using phosphate (and NaCl where required) at 0.1 M

9.2.1.3 Acetate kinase (AK) activity assays

The assay used for AK activity was a coupled assay monitoring the acetyl-phosphate dependent $NADP^+$ reduction at 340 nm. The coupled assay scheme is described below:

1. Acetate kinase

$$\text{Acetyl-phosphate} + \text{ADP} \leftrightharpoons \text{Acetate} + \text{ATP}$$

2. Hexokinase

$$\text{Glucose} + \text{MgATP}^{2-} \leftrightharpoons \text{Glucose-6-phosphate} + \text{MgADP}^- + \text{H}^+$$

3. Glucose-6-phosphate dehydrogenase

$$\text{Glucose-6-phosphate} + \text{NADP}^+ + \text{H}_2\text{O} \leftrightharpoons \text{6-Phospho-gluconate} + \text{NADPH} + \text{H}^+$$

Coupling substrates and enzymes were in excess during the assays. Although these coupling enzymes were from mesophilic sources, they were stable enough at 60°C for the rate to only be dependent on the concentration of acetate kinase. The reported AK activity measurements are the result of single enzyme assays.

Chemical	**Stock conc^n (mM)**	**Volume added (µl)**	**Assay conc^n (mM)**
HEPES buffer pH 6.5, 0.1 mM zinc acetate, 5 mM $MgCl_2$	50	830	~50
NaCl	4000	20	80
Acetyl-phosphate	100	10	1
$NADP^+$	8	50	0.4
ADP	50	20	1
Glucose	50	50	2.5
Enzyme / cell extract		5	
Coupling Enzyme	**Activity in assay (U)**		**Volume added (µl)**
Hexokinase (HK) 400 U/ml	2		5
Glucose-6-phosphate dehydrogenase (G6P-DH) 200 U/ml	2		10

Table 9-3: AK coupled assay standard conditions.

- Buffer pre-incubated at 60°C
- G6P-DH dissolved in 5 mM glycine
- Substrates and HK made up in assay buffer
- Substrates stored on ice
- Reaction started by addition of enzyme
- Ionic strength maintained at 0.1 M with NaCl

9.2.2 Fermentation run (TM444)

TM444 fermentation cell pellets used for this section of the project were produced using a Biostat CT-DCU 5-2 fermenter system (Sartorius) following TMO Renewables standard protocols. The method used is summarised in Table 9-4. The fermentation run of TM444 was carried out by the fermentation team at TMO Renewables. The analytical team determined metabolite concentrations using HPLC analysis, following standard protocols.

Fermentation parameter	Value
Media	50 mM USM + 6% Glucose + 2% YE + 2x AA Mix 1
Working volume (L)	5
Inoculum (%v/v)	500 ml (10%) of an OD_{600nm} 12.6 seed culture
Temperature (°C)	60
pH	6.7
Agitation (RPM) pre “anaerobic switch”	400
Aeration (L/min) pre “anaerobic switch”	5
Anaerobic switch (OD_{600nm})	6.3
Agitation (RPM) post “anaerobic switch”	220-350 (redox control)
Aeration (L/min) post “anaerobic switch”	1
Post switch redox control (mV)	-330

Table 9-4: TM444 fermentation parameters.

50 ml samples of the fermenter culture were taken at each time point. Cells were harvested by centrifugation (3220 x *g* for 20 min at 4°C); the supernatant was taken for HPLC analysis and the cell pellets were frozen at -20°C.

Cell extracts were produced as described in Section 2.6, and protein concentrations for each soluble cell-extract sample were determined by the Bradford assay.

9.3 Results

9.3.1 TM242 kinetic parameter determination (aldDH and PAT)

Kinetic parameters for aldDH and PAT were determined in a TM242 fermentation cell extract prepared as in Section 2.6. AldDH activities were determined using the DTNB assay (as described in Section 2.13.1.1). Substrate inhibition was not observed for aldDH under these assay conditions.

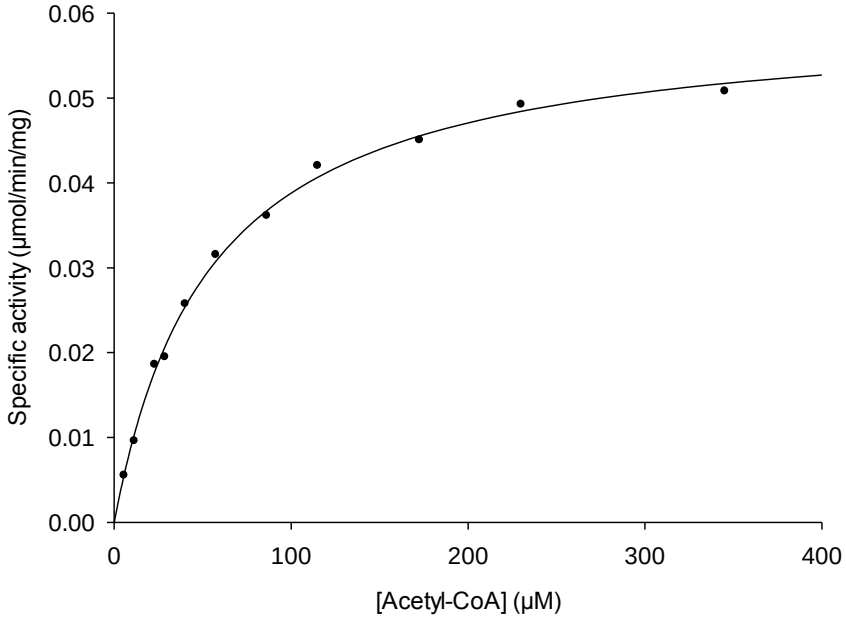


Figure 9-2: Michaelis-Menten plot of the dependence of aldDH enzymic activity (U mg^{-1} of protein) in a TM242 cell extract on the concentration of acetyl-CoA (µM) at a fixed concentration of NADH (0.18 mM).

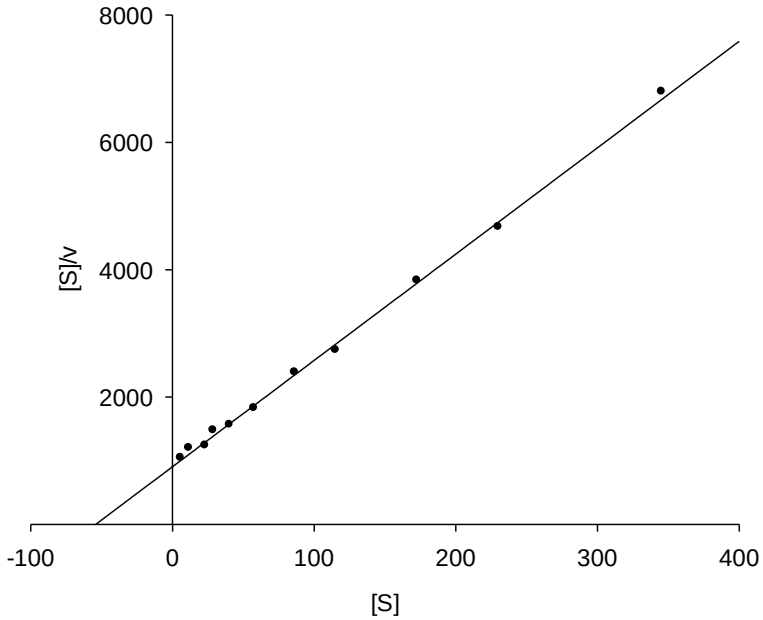


Figure 9-3: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of aldDH activity (U mg^{-1}) with respect to the concentration of acetyl-CoA (µM).

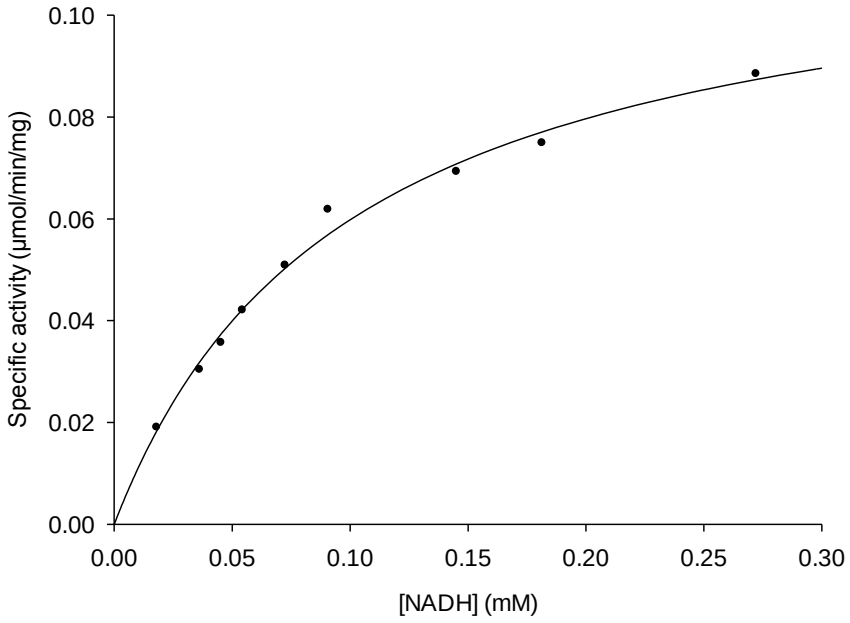


Figure 9-4: Michaelis-Menten plot of the dependence of aldDH enzymic activity (U mg^{-1} of protein) in a TM242 cell extract on the concentration of NADH (mM) at a fixed concentration of acetyl-CoA (0.46 mM).

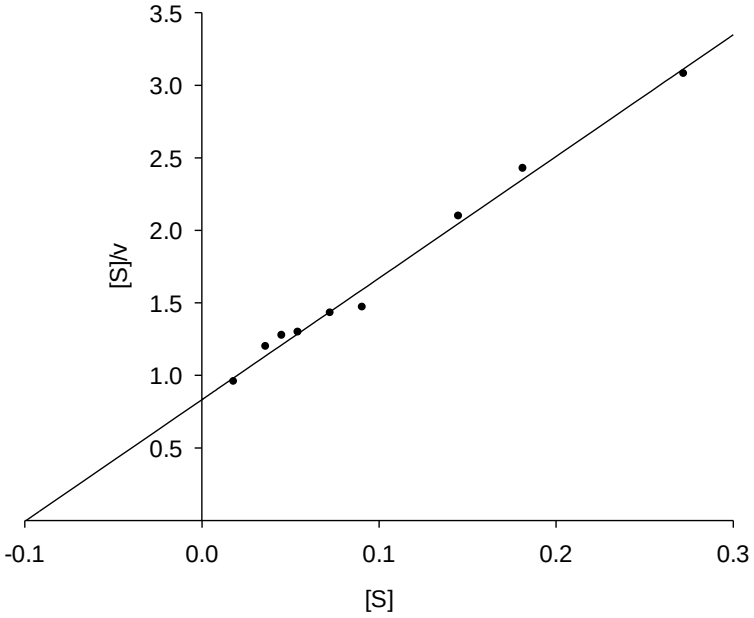


Figure 9-5: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of aldDH activity (U mg^{-1}) with respect to the concentration of NADH (mM).

PAT activities were determined using a 5 x 5 variation of the described PAT assay (Section 9.2.1.2). That is, activities were determined with 5 different concentrations of one substrate, in the presence of 5 different concentrations of the other substrate. Kinetic parameters for each substrate were then determined using secondary plots.

Activity (substrate)	K_m (mM)	Standard Error	V_{max} (U mg^{-1})	Standard Error
aldDH: (acetyl-CoA)	0.052	+/- 0.001	0.094	+/- 0.004
aldDH: (NADH)	0.104	+/- 0.004	0.134	+/- 0.004
PAT: (acetyl-CoA)	0.065	+/- 0.006	2.51	+/- 0.14
PAT: (P_i)	7.21	+/- 0.38	2.86	+/- 0.09

Table 9-5: Summary of kinetic parameters determined for aldDH and PAT in TM242 cell extracts. Assays carried out in 50 mM HEPES buffer pH 6.5, 0.1 mM zinc acetate with a maintained ionic strength of 0.1 M. V_{max} reported for aldDH was adjusted for saturation.

9.3.2 TM444 fermentation profiles

A characteristic glucose fermentation profile was observed for the TM444 fermentation run (TMO Renewables 2012 personal communication).

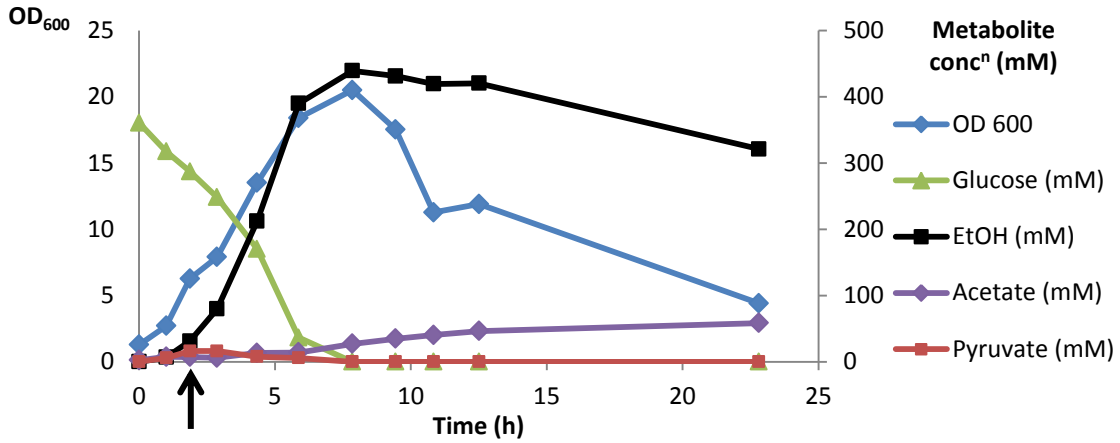


Figure 9-6: Fermentation profile of TM444 used for expression level experiments (OD_{600nm} & metabolite concentrations (mM) against time (h)). Black arrow indicates "anaerobic switch".

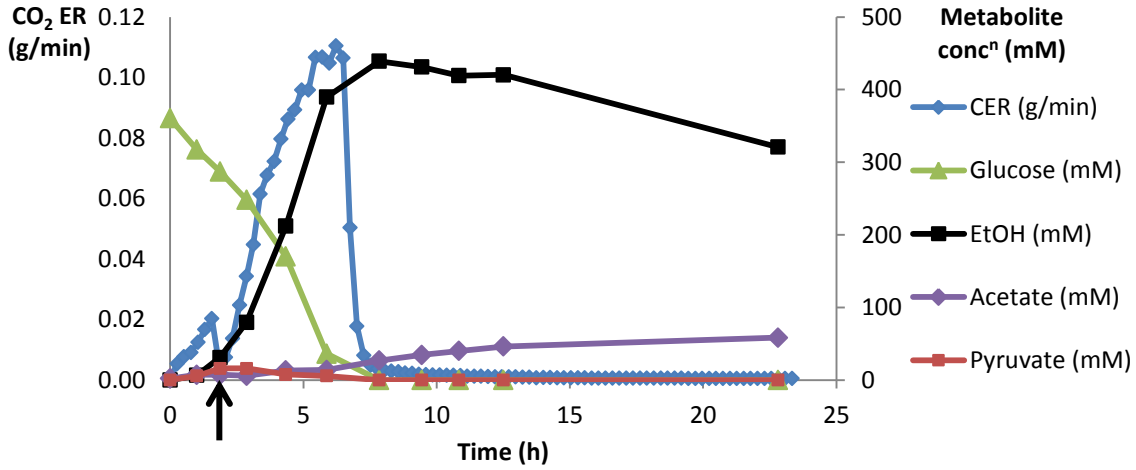


Figure 9-7: Fermentation profile of TM444 used for expression level experiments (CO_2 evolution rate (g/min) (CER) & metabolite concentrations (mM) against time (h)). Black arrow indicates "anaerobic switch".

9.3.3 Specific activity measurements

With the exception of AK, assays on each cell extract were carried out in triplicate and an average taken. Assays of PAT and AK activity were carried out on the same cell extract samples as the ADHE and PDH assays.

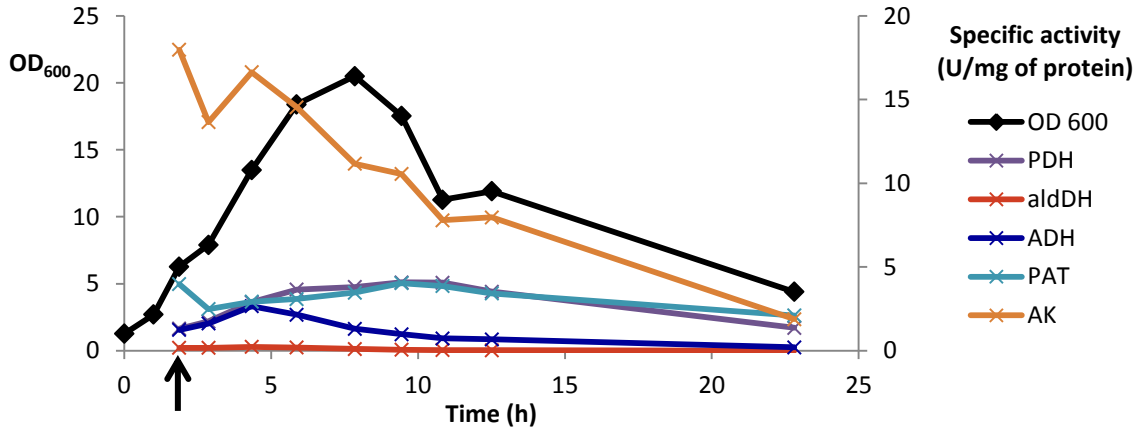


Figure 9-8: Plot of OD_{600nm} & average specific activity measurements (U mg^{-1} protein) against time (h) for the TM444 fermentation run. Black arrow indicates "anaerobic switch".

It is interesting to observe that AK appears to be the highest activity within the cell extract whereas the aldDH activity appears to be the lowest. To allow the expression pattern of each enzyme to be compared, the percentage activity of the maximum specific activity observed for each enzyme was calculated.

9.3.4 Percentage activity measurements

These transformations of the assay data show that the various enzymes assayed here are under different expression patterns.

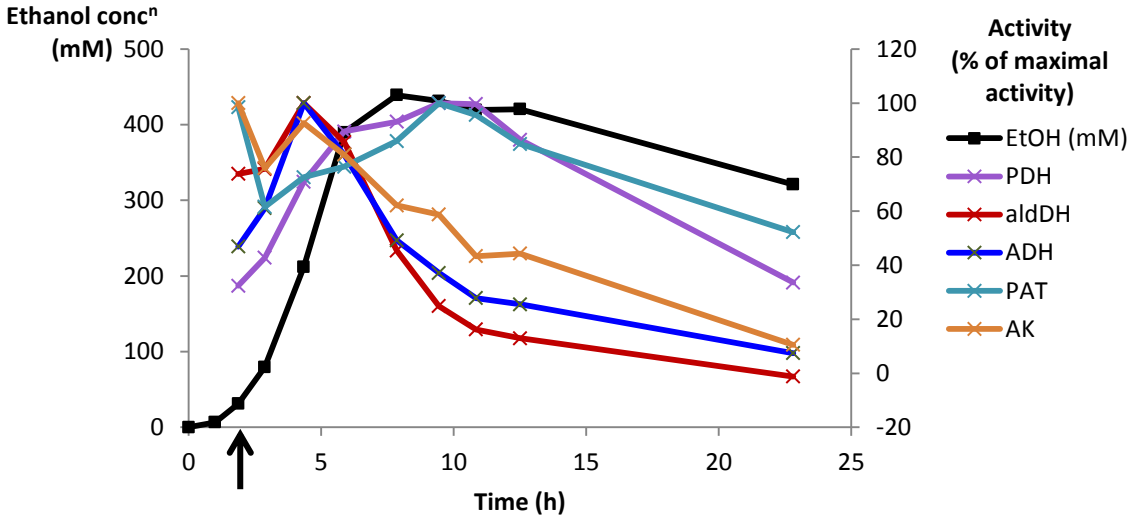


Figure 9-9: Plot of ethanol concentration (mM) and average enzyme activity measurements (% of maximum observed) against time (h) for the TM444 fermentation run. Black arrow indicates "anaerobic switch".

Results from each of the distinct areas of metabolism are visualised independently for clarity.

9.3.5 PDH expression

PDH expression appeared to increase following the switch into “anaerobic” conditions. The initial accumulation of pyruvate in the fermentation medium indicates PDH activity may be limiting pyruvate metabolism. Once PDH expression is up-regulated, this excess pyruvate appears to be metabolised coincident with a rapid increase in metabolism of glucose to ethanol. PDH expression appears to increase to a maximum and remain constant until cell numbers begin to fall.

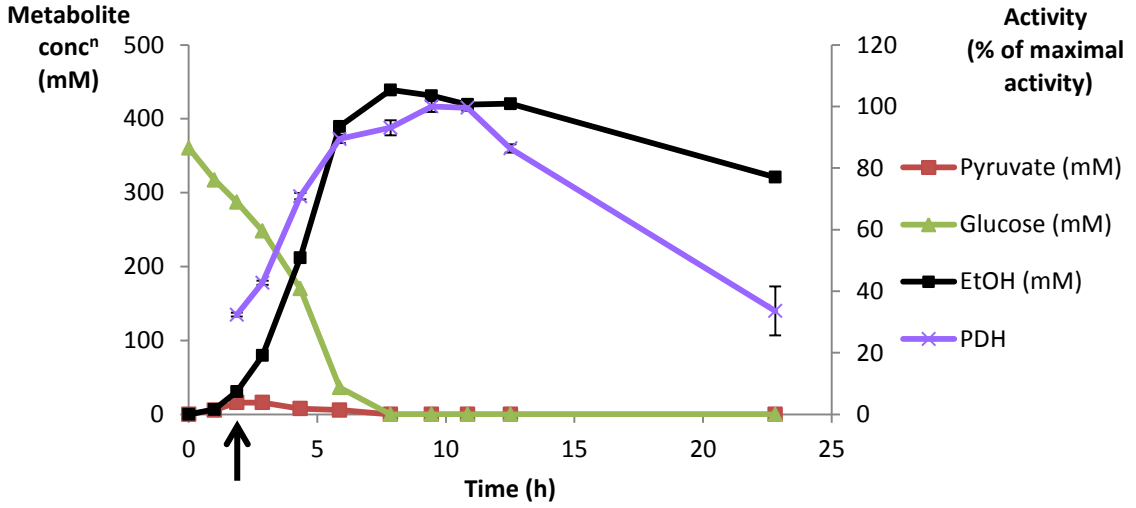


Figure 9-10: Plot of metabolite concentration (mM) and average PDH activity measurements (% of maximum observed) against time (h) for the TM444 fermentation run. Error bars indicate the standard deviation of 3 enzyme assays. Black arrow indicates “anaerobic switch”.

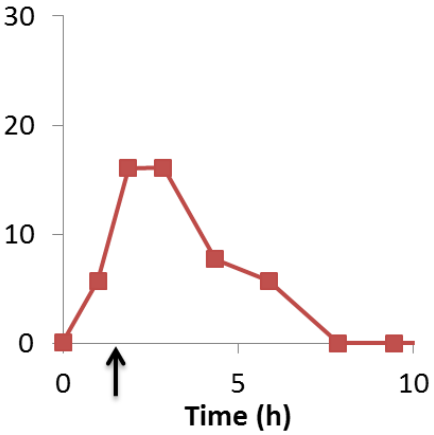


Figure 9-11: Enlarged plot of pyruvate concentration (mM) against time (h) (0-10 h) for the TM444 fermentation run. Black arrow indicates “anaerobic switch”.

9.3.6 AldDH and ADH expression

The aldDH and ADH activities associated with ADHE follow the same expression pattern and appear to be correlated with the rate of ethanol production observed for the fermentation. Unlike the other activities observed during this experiment, the expression of ADHE appears to be transient. Initial expression of ADHE increased until a majority of the glucose had been utilised, after which the activity then appeared to decrease rapidly.

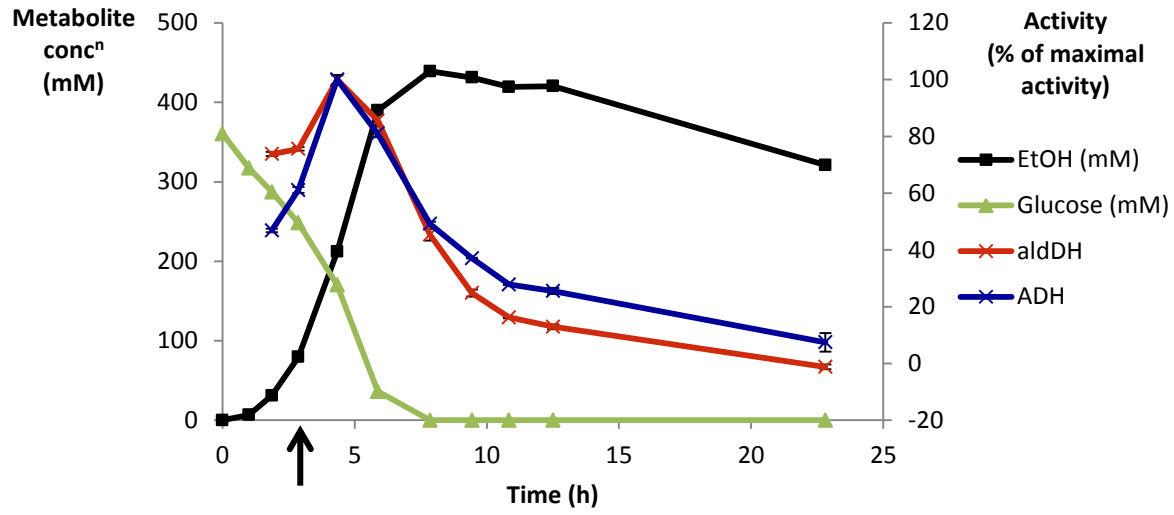


Figure 9-12: Plot of metabolite concentration (mM) and ADHE activity measurements (% of maximum observed) against time (h) for the TM444 fermentation run. Error bars indicate the standard deviation of 3 enzyme assays. Black arrow indicates "anaerobic switch".

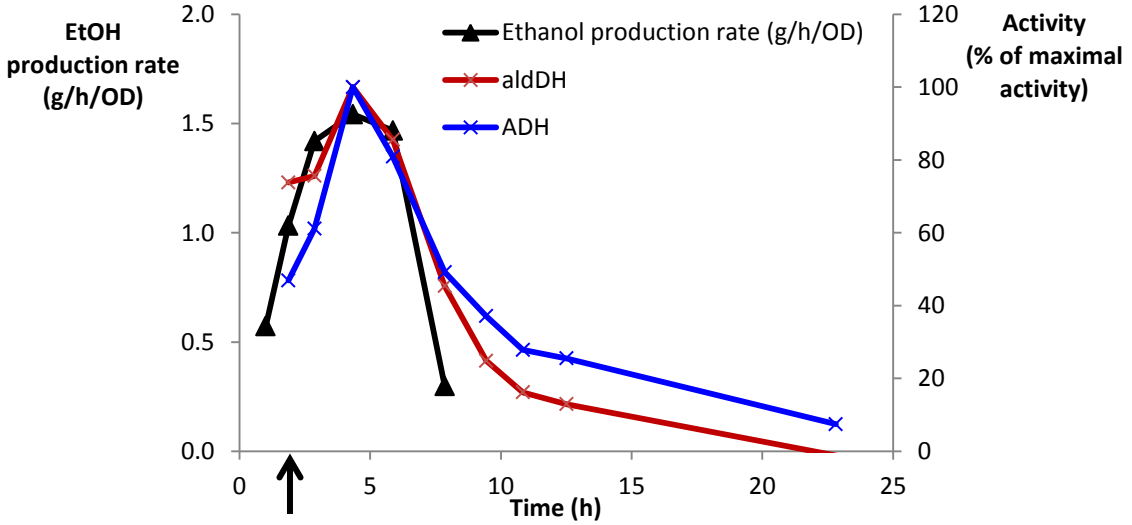


Figure 9-13: Plot of Ethanol production rate (g/h/OD) and ADHE activity measurements (% of maximum observed) against time (h) for the TM444 fermentation run. Black arrow indicates "anaerobic switch".

9.3.7 PAT and AK expression

PAT and AK expression appears to be influenced less by the switch to anaerobic conditions than the other enzymes monitored. Acetate production through the fermentation appears to partially reflect the expression of PAT. Significant acetate accumulation does not occur until a majority of the glucose has been utilised.

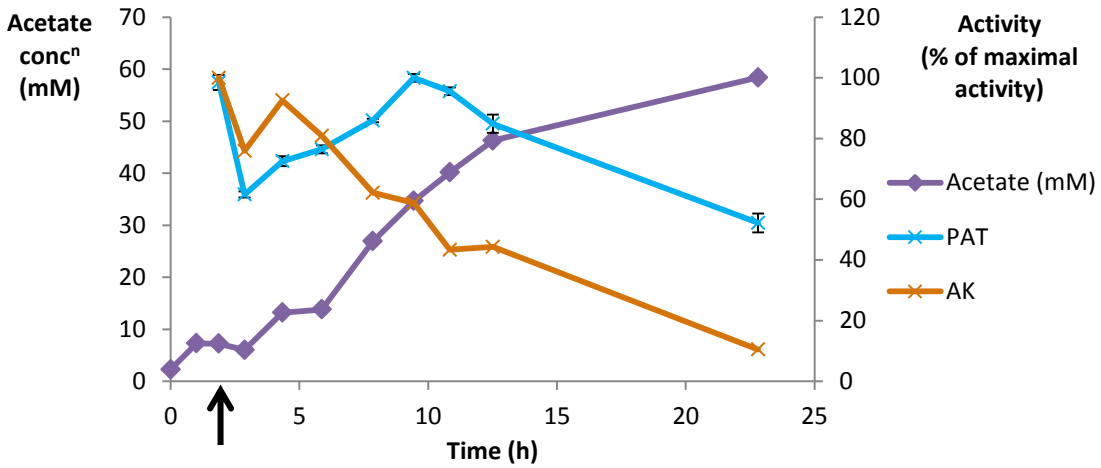


Figure 9-14: Plot of metabolite concentration (mM), PAT & AK activity measurements (% of maximum observed) against time (h) for the TM444 fermentation run. Error bars indicate the standard deviation of 3 enzyme assays for PAT activity measurements. Black arrow indicates “anaerobic switch”.

9.4 Discussion

Experiments carried out here aimed to determine the expression profiles of various key enzymes through a fermentation run. The kinetic properties of two acetyl-CoA consuming enzymes (PAT and aldDH) were also determined under identical conditions.

As discussed in the introduction to this chapter, the TM242 and TM444 strains have been shown to metabolise glucose to ethanol efficiently under fermentative conditions; yields of ethanol can approach 90% of the theoretical maximum (0.51 g ethanol /g glucose) (Cripps et al. 2009). It is important to develop an understanding of the enzymes responsible for producing ethanol or other by-products during this process.

Acetyl-CoA metabolism is a potential bottle-neck in ethanol production (as illustrated by the accumulation of pyruvate observed during these experiments), two relevant competing activities act on this metabolite under fermentative conditions. Under these conditions a majority of the carbon flux is assumed to be towards ethanol, as this is the major product seen during fermentation. The possibility of redundant cycling via the

hypothesised ACS although unlikely to be physiologically relevant should also be considered. Enzyme expression levels, kinetic properties, metabolite concentrations and the presence of other regulatory factors may all play a role in directing the carbon flow.

The activity of PAT within cell extracts of TM242 is consistently higher than that observed for the aldDH domain of ADHE. Kinetic characterisation of these two activities under near identical conditions demonstrated that the V_{max} of PAT (2.5 U mg^{-1}) is significantly higher than for the aldDH domain of ADHE (0.09 U mg^{-1}). The relatively high K_m observed for PAT with respect to phosphate suggests this substrate may play a role in regulating enzyme activity *in vivo,* as the K_m values observed for acetyl-CoA for the two proteins are similar. This is further illustrated by the significant difference between the V_{max}/K_m values for the aldDH and PAT activities. As described in Chapter 3, V_{max}/K_m values should be treated with caution as the rate obtained *in vivo* will be dependent on physiological substrate concentrations.

Activity	Substrate	V_{max} (U mg^{-1})	K_m (mM)	V_{max}/K_m ((U mg^{-1})/mM)
TM242 cell extract ADHE				
aldDH	Ac-CoA	0.09	0.05	1.8
aldDH	NADH	0.13	0.10	1.3
TM242 cell extract PAT				
PAT	Ac-CoA	2.50	0.07	38.6
PAT	Pi	2.90	7.21	0.4

Table 9-6: Kinetic parameters for acetyl-CoA consuming activities in TM242 cell extract.

Control of metabolic flux to ethanol observed in TM444 cells may be regulated by free phosphate concentrations. Phosphate has been suggested as an effector of metabolic enzymes (Goel et al. 2012). For example, it has been shown using NMR that phosphate concentration levels vary significantly throughout metabolism in *Lactococcus lactis* (between 5 to 35 mM) in response to glucose levels (Neves et al. 2002). In *Zymomonas mobilis* fermentations on glucose, the intracellular free phosphate concentration was determined by NMR to be 0.5 mM (De Graaf et al. 1999). If the intracellular phosphate concentration was around this figure for TM444, PAT would be operating well under its maximum velocity. It can be conceived that the variation in phosphate concentration within cells may play a key role in the regulation of PAT activity. However, without determining the intracellular concentration of

phosphate within TM242/TM444 cells through a fermentation run, it is not possible to confirm this hypothesis.

Substrate inhibition was not observed for the aldDH activity of ADHE under these assay conditions. The DTNB assay reflects the NADH-dependent CoA-SH release, corresponding to the aldDH activity in isolation. Inhibition was observed previously for the aldDH domain using the NADH assay. As this assay was shown to be a measure of metabolic flux to ethanol, inhibition may have been caused by limitations in the flux of acetaldehyde to the ADH domain of the protein. As the DTNB assay cannot be carried out at the pH of the normal assay buffer used for ADHE, it was not confirmed whether the lack of inhibition was due to pH/buffer effects or effects of co-assaying with the ADH domain.

Assays of enzyme activities through a TM444 fermentation run were carried out. It should be noted that although samples were taken over a 23 h period, the production phase of the fermentation is considered complete at the point of maximal ethanol yield (at approximately 8 h). Specific activities, which are based on protein concentration, were used in an effort to allow enzyme activities independent of cell number to be quantified. The peaks of enzyme activities appear to be independent of OD_{600nm}; this implies that activities reported have been successfully resolved from cell number. There are limitations to these assays but the results serve as an indication as to where the peak of enzyme expression lies for each activity. As different expression patterns were observed for the activities monitored through the fermentation, it is suggested that varied mechanisms of gene regulation may be involved for the enzymes monitored.

Steady-state assays cannot be carried out under truly physiological conditions, and they were not fully optimised for each of the activities measured; therefore comparing the specific activities directly would be inappropriate. It is an interesting observation that AK appeared to be the most active enzyme measured, whereas the aldDH activity of ADHE appeared to be the lowest. If these assays were a true reflection of expression levels, the intracellular concentrations of phosphate and NADH are likely to be crucial in regulating metabolic flux to ethanol.

The production of acetyl-CoA is regulated by the PDH protein complex. The expression of PDH has been up-regulated by placing the operon under the control of an anaerobic LDH promoter. The expression pattern observed here is consistent with anaerobically regulated expression. The initial pyruvate accumulation within the

fermentation medium is likely to be caused by a shortfall in PDH activity. This was quickly overcome following the anaerobic switch due to induction of PDH expression. Once expression of PDH had been induced, the activity increased to a maximum level where it remained constant prior to the loss of activity due to the initiation of cell death.

AldDH and ADH activities observed within the cell extract followed the maximal rate of ethanol production as would be predicted. As discussed in the Introduction to this thesis, a REX repressor protein target site has been identified within the *adhE* promoter region. The anaerobic induction of expression is consistent with the suggested REX control over the promoter. The protein in *Bacillus subtilis* has been proposed to bind to operator regions of DNA within the promoter of the target gene. This binding is linked to the redox balance within the cell through NADH/NAD^+ interactions. When NADH levels are low, REX can interact with the DNA and repress expression; when NADH levels are high, REX interacts with NADH and is released from the DNA target region, thereby removing repression (Wang et al. 2008).

Once a majority of the glucose had been metabolised during the fermentation run, the ADHE activities rapidly decreased. This transient activity observed for ADHE is unusual and suggests that this protein may be rapidly turned-over within the cell. This decrease is coincident with the rapid decrease in CO_2 evolution rate when the carbon source has been exhausted. There may also be a fall in NADH levels as glycolysis is reduced due to glucose depletion, hence a shutdown in expression of REX-regulated genes. As no aldDH-independent peak of ADH activity was observed during these experiments, it would appear that the hypothesised “ADH2” activity (described in Chapter 3) mirrors ADHE expression.

Acetate production appeared to follow the PAT expression pattern observed through the fermentation. A majority of the acetate accumulated once glucose had been exhausted, and was coupled with a rapid reduction in CO_2 evolution rate, all of which occurred outside the production phase of fermentation. This increase in acetate may be a combination of cell lysis (which begins to occur once the carbon source has been fully utilised) and the secondary metabolism of ethanol. AK is highly expressed during both the aerobic and anaerobic stages of growth. PAT expression appears to occur during both aerobic and anaerobic phases, but expression levels increased as the anaerobic phase progressed. The expression patterns observed indicate that it may be metabolite concentrations, rather than enzyme expression levels, which control metabolic flux to ethanol instead of acetate. The essential regeneration of NAD^+ from

NADH, as well as the intracellular concentration of phosphate, are both likely regulatory factors of the PAT – aldDH metabolic competition.

Attempts to produce viable PAT and AK knock-out strains have been unsuccessful (TMO Renewables 2012 personal communication) suggesting these activities are essential for cell survival. This requirement is thought to be due to the essential regeneration of ATP. The low levels of acetate produced initially during glucose fermentation would suggest this is not a major metabolic conversion. Attempts are continuing to produce PAT/AK knock-out strains to observe the effect on metabolism. If such a strain cannot be obtained, attempts to down-regulate these enzymes through promoter switching may be tested.

In conclusion, under the assay conditions used here, the maximum specific activity of PAT appears to be significantly higher than that of the aldDH activity of ADHE. Values of K_m for the enzymes suggest that a possible regulator of metabolic flux to ethanol is the intracellular concentrations of phosphate and NADH. As the intracellular concentrations of phosphate and NADH in TM242/TM444 are yet to be determined, this hypothesis remains untested. Expression of both ADHE and PDH appear to be under anaerobic regulation, whereas AK and PAT expression levels show little variation between aerobic and anaerobic growth conditions. More information may be obtained through a similar set of experiments on either fed-batch or continuous-fermentation cell samples. Such experiments would provide further information regarding the expression patterns of the enzymes, particularly the transient expression of ADHE. Enzyme expression patterns during a fermentation where higher levels of acetate accumulate (e.g. xylose as carbon source (TMO Renewables 2012 personal communication)) may also prove informative.

10 GENERAL DISCUSSION AND FUTURE WORK

Bioethanol produced from ligno-cellulosic waste materials has the potential to have a significant impact on the global dependence on petrol as a transportation fuel. Clearly a range of different technologies will be required to meet current energy requirements free from fossil fuels. Without the development of such technologies the environment in which we reside is likely to be adversely affected during future generations. The organism *Geobacillus thermoglucosidasius* TM242 has the potential to offer commercially-viable ethanol production on a large scale, utilising a range of sustainable feedstocks. Further manipulation of this organism to optimise ethanol yields could substantially improve the technology currently under development by TMO Renewables Ltd. The ADHE protein that has been the subject of this investigation has been shown to be a key enzyme required for the production of ethanol by the TM242 organism. The work presented here seeks to further the understanding of this key protein and aid the optimisation of ethanol production in the TMO Renewables process.

Ethanol production has been shown to be dependent on the presence of the *adhE* gene product, despite an additional 7 genes being annotated in the TM242 genome as alcohol dehydrogenases (ADHs). A second ADH activity ("ADH2") within cell extracts of TM242 was detected here, but the gene coding for the responsible protein remains to be determined. The possibility of ADHE presenting the major aldDH activity *in vivo*, with other proteins contributing further ADH activities, cannot be ignored. The role of these proteins in metabolism could be demonstrated using gene knock-out experiments, followed by strain characterisation. Cloning and recombinant expression of these genes followed by kinetic characterisation may also prove informative.

The relationship between ADHE and the PAT-AK branches of metabolism has been investigated through preliminary studies carried out here. Coincident with the induction of fermentative metabolism, increases in the enzyme activities associated with ADHE in cell extracts (indicative of gene expression) were observed, but the PAT and AK activities appeared to vary much less significantly. This type of examination will be carried out on samples from a continuous or fed batch fermentation system, to investigate further the balance of metabolic activities during ethanol production. Differences in enzyme activity levels during growth on different substrates are also of particular interest. The specific activities of the competing enzymes *in vitro* suggest that the direction of metabolic flux may be heavily influenced by the intracellular

concentrations of phosphate and NADH, and therefore quantification of the concentrations of these key metabolites will prove informative. Techniques such as ^{31}P-NMR may be used to determine the levels of intracellular phosphate (De Graaf et al. 1999), whereas adapted methods developed for other organisms, e.g. extraction followed by HPLC separation and UV-Vis absorbance detection (Sporty et al. 2008), may allow concentrations of NAD^+ and NADH to be determined. Attempts to knock-out or down-regulate the PAT and AK genes are on-going, as this branch of metabolism is in direct competition with ethanol production. It currently seems unlikely that viable knock-out strains can be produced, but determining the minimum level of gene expression through promoter switching may enable further optimisation. The effect of up-regulating the ADHE gene expression through the use of a stronger promoter sequence is also currently under investigation.

The characterisation of ADHE and PAT provides significant insight into the catalytic activity of the proteins; characterisation of the activities under native conditions would provide further insight into the *in vivo* activity of these enzymes. Purification and assay of the ADHE protein under anaerobic conditions would allow the aldDH activity of the protein to be resolved free from the aerobic loss of activity observed for this domain. It is hypothesised that oxidation of the catalytic cysteine (Cys_{257}) described during the Introduction, may be the cause of the instability through purification of this protein.

Although the experiments performed as part of Chapter 9 go some way to elucidating the expression pattern of ADHE in relation to other key metabolic enzymes, evaluation of gene expression patterns through an ethanol-generating fermentation may be more comprehensively resolved using RNA analysis techniques. This will form part of a recently commenced project at The University of Bath in conjunction with TMO Renewables, where RNA-seq techniques will be used to identify and monitor all mRNAs produced under various culture conditions. These studies will provide valuable information concerning the influence of gene expression on the regulation of metabolic flux within the fermenting organism.

The redox sensing protein REX is a likely regulator of gene expression of *adhE* (Paget, M. S. 2012 unpublished work). This regulation of genes will also be a significant focus of future research to illuminate the mechanisms involved in fermentative gene expression.

The ADH domain of the ADHE protein has been structurally resolved independently of the aldDH activity; this is the first reported high-resolution structure of this domain of an ADHE. Several loop regions show significant differences compared to those observed in homologous structures within the PDB, and it is hypothesised that interactions with the aldDH domain may be coordinated through some of these loops. *In silico* modelling work has predicted a potential interaction between the two domains of ADHE, such that spirosome-like assemblies could form. This is consistent with the large M_r complexes observed for ADHE during Chapter 3. The relatively high K_m consistently observed for the ADH domain of ADHE for acetaldehyde might be alleviated by substrate channelling between the two ADHE domains. This would prevent the accumulation of the potentially toxic intermediate within the cell, and maximise the efficiency of NAD^+ regeneration under anaerobic conditions. *In silico* modelling work has not currently been able to confirm this hypothesis due to the significant distance predicted between active sites of the two domains. The formation of the spirosome complexes observed for ADHE proteins may also influence the mechanism of substrate channelling.

The resolved structure of the ADH domain is already providing a valuable resource for guiding the manipulation of the ADHE protein activity through site-directed mutagenesis. The phenotypic effect of the point mutations described by Brown et al. (2011) and Shao et al. (2011) in *Clostridium thermocellum* are currently being investigated in the ADHE of *G. thermoglucosidasius*. Modifications of other key residues validated using the solved ADH domain structure are also being investigated to observe the effect of modifying the ADH activity of ADHE. The effect on enzyme activities as well as phenotypic effects in fermentation experiments will be monitored, to determine the biochemical basis of any industrially significant changes observed. Some promising results have already been obtained and are currently being fully evaluated (TMO Renewables. 2012 personal communication). An example of such a targeted residue is Asp_{558}, which was shown to be potentially involved in hydrogen bonding of the nicotinamide ring. The effect of removing the steric and electrostatic clash of the Asp_{493} of ADHE (Asp_{57} in Fragment 11) with the additional phosphate group of $NADP^+$, is also being evaluated to determine the effect on nicotinamide cofactor preference, and potential implications thereof for fermentative metabolism. These experiments will be focussed initially around “loss of charge” mutations, replacing the residue with a similarly sized un-charged residue such as asparagine, or “loss of bulk” mutations by replacing with significantly smaller residues such as glycine.

Although X-ray crystallography has shown a coordinated metal ion within the active site of the ADH domain, the physiologically relevant ion has not been unambiguously determined here. Activity was observed in the absence of additional metal ions, but stimulation of activity with several metal ions was also shown in cell extracts; however, this was not the case for the purified recombinant proteins studied here. It is hypothesised that the use of metal-affinity purification for these proteins may have allowed scavenging of additional metal ions during that purification, thus saturating the recombinant ADH domains and preventing further stimulation through the addition of metal ions to the assays. Zn^{2+} was used for most of the assays during this project whereas Fe^{2+} has previously been suggested to be a key cofactor for ADHE proteins (Kessler et al. 1992). The ADH domain also shows significant sequence homology to the Fe^{2+} containing ADH superfamily. To resolve the physiologically relevant ion, the fully-purified native ADHE protein should be subject to ion analysis experiments. As the native ADHE protein was not fully resolved during this work, such experiments could not be carried out during the experimental phase of the project.

Despite several different sized aldDH coding fragments being generated during this project, the aldDH activity of ADHE could not be resolved independently of the ADH domain. Other groups have achieved this isolation for the ADHE of *Entamoeba histolytica* (Chen et al. 2004), but similar fragmentation carried out here did not yield active protein.

Structural resolution of the full-length ADHE protein remains a priority. Information on the formation of the large multimeric structures observed, any potential channelling of acetaldehyde between the aldDH and ADH domains, and identification of key residues that may be manipulated to modulate the activities of the ADHE protein may be inferred from such a structure. If crystal formation remains elusive, optimisation of high-resolution electron microscopy techniques, or Small Angle X-ray Scattering (SAXS) methods, may provide further structural insights into the nature of the interactions between the domains of ADHE. Disruption of interacting residues predicted to be involved in multimeric assembly through site-directed mutagenesis, may allow disruption of the spirosome structures, and potentially allow a high-resolution crystal structure to be determined for the ADHE protein in isolation. The flexible nature of the linker between the two domains of ADHE may however persist in hindering crystal formation.

Although the AcAldDH protein identified and characterised here shows a high catalytic activity, it appears that product inhibition of this protein may limit its current usefulness in terms of maximising metabolic flux to ethanol in TM242. The high-resolution crystal structure determined for this protein may provide sufficient insight to manipulate the activity to overcome this product inhibition. Structurally-informed site-directed mutagenesis has been used previously to overcome inhibition of enzymes (Hu et al. 2010; Kai et al. 2006), although investigation of this was not possible in the timescale of this project. Further optimisation of co-crystallisation conditions may allow a higher resolution structure for the AcAldDH protein (with substrate/product bound) to be obtained. Such a structure could better inform the mutagenesis targeting required to overcome the product inhibition observed.

The successful creation of stable artificial ADHE proteins (Fusions) was demonstrated during this project. The catalytic effectiveness of these Fusion proteins appeared limited, although experiments to evaluate these proteins gave contrasting results. Evaluation of the performance of the Fusion 1 protein in a sporulation-deficient *G. thermoglucosidasius* strain would allow assessment of the hypothesised difference between tube fermentation and full fermentation conditions. Successful optimisation of the AcAldDH protein activity through site-directed mutagenesis may yield an improved Fusion component domain that may demonstrate improved activity *in vivo*.

The ADHE enzyme studied here is an essential enzyme in ethanol production in *G. thermoglucosidasius*. The characteristics of ADHE and its component domains that have been reported in this thesis are now being studied further, and hopefully exploited, at TMO Renewables where the role of this protein in terms of the *rate* of ethanol production and *tolerance* of the organism to ethanol are of particular interest.

11 REFERENCES

Angers-Loustau, A., Rainy, J., & Wartiovaara, K. (2007). PlasmaDNA: a free, cross-platform plasmid manipulation program for molecular biology laboratories. *BMC Mol.Biol.* **8**, 77

Antoni, D., Zverlov, V.V., & Schwarz, W.H. (2007). Biofuels from microbes. *Appl.Microbiol.Biotechnol.* **77**, 23-35

AppliChem. Bilogical buffers.
http://www.applichem.com/fileadmin/Broschueren/BioBuffer.pdf

Arnau, J., Jorgensen, F., Madsen, S.M., Vrang, A., & Israelsen, H. (1998). Cloning of the *Lactococcus lactis adhE* gene, encoding a multifunctional alcohol dehydrogenase, by complementation of a fermentative mutant of *Escherichia coli*. *J.Bacteriol.* **180**, 3049-3055

Asanuma, N., Yoshii, T., & Hino, T. (2004). Molecular characteristics and transcription of the gene encoding a multifunctional alcohol dehydrogenase in relation to the deactivation of pyruvate formate-lyase in the ruminal bacterium *Streptococcus bovis*. *Arch.Microbiol.* **181**, 122-128

Atteia, A., van, L.R., Mendoza-Hernandez, G., Henze, K., Martin, W., Riveros-Rosas, H., & Gonzalez-Halphen, D. (2003). Bifunctional aldehyde/alcohol dehydrogenase (ADHE) in chlorophyte algal mitochondria. *Plant Mol.Biol.* **53**, 175-188

Barley, S., 2-9-2010. Where there's bugs, there's brass: British firm lands $500m biofuel contract, *The Guardian*, 26.

Barnard, D., Casanueva, A., Tuffin, M., & Cowan, D. (2010). Extremophiles in biofuel synthesis. *Environ.Technol.* **31**, 871-888

Bartosiak-Jentys, J. 2010. Metabolic engineering and Metabolic Flux Analysis of thermophilic, ethanologenic *Geobacillus* spp. (PhD Thesis). Imperial College, University of London.

Botting C.H. http://www.st-andrews.ac.uk/~bmsmspf/PDFs/Publishing.pdf

Bradford, M.M. (1976). Rapid and Sensitive Method for Quantitation of Microgram Quantities of Protein Utilizing Principle of Protein-Dye Binding. *Anal.Biochem.* **72**, 248-254

Brekasis, D. & Paget, M.S. (2003). A novel sensor of NADH/NAD^+ redox poise in *Streptomyces coelicolor* A3(2). *EMBO.J.* **22**, 4856-4865

BRENDA enzyme database. BRENDA enzyme database.
http://www.brenda-enzymes.org/php/result_flat.php4?ecno=1.1.1.1

Brown, S.D., Guss, A.M., Karpinets, T.V., Parks, J.M., Smolin, N., Yang, S., Land, M.L., Klingeman, D.M., Bhandiwad, A., Rodriguez, M., Jr., Raman, B., Shao, X., Mielenz, J.R., Smith, J.C., Keller, M., & Lynd, L.R. (2011). Mutant alcohol dehydrogenase leads to improved ethanol tolerance in *Clostridium thermocellum*. *Proc.Natl.Acad.Sci.U.S.A* **108**, 13752-13757

Bruchhaus, I. & Tannich, E. (1994). Purification and molecular characterization of the NAD(+)-dependent acetaldehyde/alcohol dehydrogenase from *Entamoeba histolytica*. *Biochem.J.* **303**, 743-748

Burdette, D. & Zeikus, J.G. (1994). Purification of acetaldehyde dehydrogenase and alcohol dehydrogenases from *Thermoanaerobacter ethanolicus* 39E and characterization of the secondary-alcohol dehydrogenase (2 ºAdh) as a bifunctional alcohol dehydrogenase--acetyl-CoA reductive thioesterase. *Biochem.J.* **302**, 163-170

Carugo, O. & Argos, P. (1997). NADP-dependent enzymes. I: Conserved stereochemistry of cofactor binding. *Proteins.* **28**, 10-28

Ceccarelli, E.A., Carrillo, N., & Roveri, O.A. (2008). Efficiency function for comparing catalytic competence. *Trends Biotechnol.* **26**, 117-118

Charles, D. (2009). Biofuels. Corn-based ethanol flunks key test. *Science.* **324**, 587

Chen, M., Li, E., & Stanley, S.L., Jr. (2004). Structural analysis of the acetaldehyde dehydrogenase activity of *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2), a member of the ADHE enzyme family. *Mol.Biochem.Parasitol.* **137**, 201-205

Chen, R. & Weng, Z. (2002). Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins.* **47**, 281-294

Chen, V.B., Arendall, W.B., III, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., & Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr.D.Biol. Crystallogr.* **66**, 12-21

Chen, Y.M. & Lin, E.C. (1991). Regulation of the *adhE* gene, which encodes ethanol dehydrogenase in *Escherichia coli*. *J.Bacteriol.* **173**, 8009-8013

Comeau, S.R., Gatchell, D.W., Vajda, S., & Camacho, C.J. (2004). ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* **32**, W96-W99

Cornish-Bowden, A. & Cardenas, M.L. (2010). Specificity of non-Michaelis-Menten enzymes: necessary information for analyzing metabolic pathways. *J.Phys.Chem.B.* **114**, 16209-16213

Cripps, R.E., Eley, K., Leak, D.J., Rudd, B., Taylor, M., Todd, M., Boakes, S., Martin, S., & Atkinson, T. (2009). Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metab.Eng.* **11**, 398-408

Dailly, Y., Mat-Jan, F., & Clark, D.P. (2001). Novel alcohol dehydrogenase activity in a mutant of *Salmonella* able to use ethanol as sole carbon source. *FEMS Microbiol.Lett.* **201**, 41-45

Dailly, Y.P., Bunch, P., & Clark, D.P. (2000). Comparison of the fermentative alcohol dehydrogenases of *Salmonella typhimurium* and *Escherichia coli*. *Microbios.* **103**, 179-196

Dan, M. & Wang, C.C. (2000). Role of alcohol dehydrogenase E (ADHE) in the energy metabolism of *Giardia lamblia*. *Mol.Biochem.Parasitol.* **109**, 25-36

Danson, M. J. 2011. 4 South 1.35, Biology and Biochemistry, University of Bath, Bath. BA2 7AY. Personal Communication

De Graaf, A.A., Striegel, K., Wittig, R.M., Laufer, B., Schmitz, G., Wiechert, W., Sprenger, G.A., & Sahm, H. (1999). Metabolic state of *Zymomonas mobilis* in glucose-, fructose-, and xylose-fed continuous cultures as analysed by ^{13}C- and ^{31}P-NMR spectroscopy. *Arch.Microbiol.* **171**, 371-385

Department for Transport. (14-5-2012). Biofuels use and supply. http://www.dft.gov.uk/topics/sustainable/biofuels/use-supply

Department for Transport. http://assets.dft.gov.uk/statistics/releases/transport-statistics-great-britain-2011/energy-summary.pdf

Department for Transport. (9-5-2012). Vehicle licensing statistics 2011. http://www.dft.gov.uk/statistics/releases/vehicle-licensing-statistics-2011/

Dokmanic, I., Sikic, M., & Tomic, S. (2008). Metals in proteins: correlation between the metal-ion type, coordination number and the amino-acid residues involved in the coordination. *Acta Crystallogr.D.Biol.Crystallogr.* **64**, 257-263

Eisenthal, R. & Cornish-Bowden, A. (1974). The direct linear plot. A new graphical procedure for estimating enzyme kinetic parameters. *Biochem.J.* **139**, 715-720

Eisenthal, R., Danson, M.J., & Hough, D.W. (2007). Catalytic efficiency and k_{cat}/K_m: a useful comparator? *Trends Biotechnol.* **25**, 247-249

Ellman, G.L. (1959). Tissue sulfhydryl groups. *Arch.Biochem.Biophys.* **82**, 70-77

Emsley, P., Lohkamp, B., Scott, W.G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr.D.Biol.Crystallogr.* **66**, 486-501

Espinosa, A. (2001). The Bifunctional *Entamoeba histolytica* Alcohol Dehydrogenase 2 (EhADH2) Protein Is Necessary for Amebic Growth and Survival and Requires an Intact C-terminal Domain for Both Alcohol Dehydrogenase and Acetaldehyde Dehydrogenase Activity. *J.Biol.Chem.* **276**, 20136-20143

Espinosa, A., Perdrizet, G., Paz, Y.M., Lanfranchi, R., & Phay, M. (2009). Effects of iron depletion on *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2) and trophozoite growth: implications for antiamoebic therapy. *J.Antimicrob.Chemother.* **63**, 675-678

Evers, T.H., Appelhof, M.A., Meijer, E.W., & Merkx, M. (2008). His-tags as Zn(II) binding motifs in a protein-based fluorescent sensor. *Protein Eng.Des.Sel.* **21**, 529-536

Fargione, J., Hill, J., Tilman, D., Polasky, S., & Hawthorne, P. (2008). Land clearing and the biofuel carbon debt. *Science.* **319**, 1235-1238

Filipe, V., Hawe, A., & Jiskoot, W. (2010). Critical evaluation of Nanoparticle Tracking Analysis (NTA) by NanoSight for the measurement of nanoparticles and protein aggregates. *Pharm.Res.* **27**, 796-810

Fontaine, L., Meynial-Salles, I., Girbal, L., Yang, X., Croux, C., & Soucaille, P. (2002). Molecular characterization and transcriptional analysis of *adhE2*, the gene encoding the NADH-dependent aldehyde/alcohol dehydrogenase responsible for butanol production in alcohologenic cultures of *Clostridium acetobutylicum* ATCC 824. *J.Bacteriol.* **184**, 821-830

Food and Agriculture Organization of the United Nations. (14-5-2012). How to Feed the World in 2050. http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf

Fox, R.J. & Clay, M.D. (2009). Catalytic effectiveness, a measure of enzyme proficiency for industrial applications. *Trends Biotechnol.* **27**, 137-140

Fuel Flex International TM. http://www.fullflexint.com/

Garman, E.F. & Owen, R.L. (2006). Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr.D.Biol.Crystallogr.* **62**, 32-47

Garsin, D.A. (2010). Ethanolamine utilization in bacterial pathogens: roles and regulation. *Nat.Rev.Microbiol.* **8**, 290-295

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R.D., & Bairoch A. (2005). *The Proteomics Protocols* Handbook*, Protein Identification and Analysis Tools on the ExPASy Server.* Humana Press.

Goel, A., Santos, F., Vos, W.M., Teusink, B., & Molenaar, D. (2012). Standardized assay medium to measure *Lactococcus lactis* enzyme activities while mimicking intracellular conditions. *Appl.Environ.Microbiol.* **78**, 134-143

Goodlove, P.E., Cunningham, P.R., Parker, J., & Clark, D.P. (1989). Cloning and sequence analysis of the fermentative alcohol-dehydrogenase-encoding gene of *Escherichia coli*. *Gene.* **85**, 209-214

Hill, J., Nelson, E., Tilman, D., Polasky, S., & Tiffany, D. (2006). Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proc.Natl.Acad.Sci.U.S.A* **103**, 11206-11210

Hills, C. 2011. Dept Biology and Biochemistry, University of Bath, Bath. BA2 7AY. Unpublished Data quoted with permission.

Holm, L. & Park, J. (2000). DaliLite workbench for protein structure comparison. *Bioinformatics.* **16**, 566-567

Hu, X., Robin, S., O'Connell, S., Walsh, G., & Wall, J.G. (2010). Engineering of a fungal beta-galactosidase to remove product inhibition by galactose. A*ppl.Microbiol. Biotechnol.* **87**, 1773-1782

IEA (2011), *Key world energy statistics 2011*.

Ingram, L.O. (1990). Ethanol tolerance in bacteria. *Crit Rev.Biotechnol.* **9**, 305-319

Ingram, L.O., Vreeland, N.S., & Eaton, L.C. (1980). Alcohol tolerance in *Escherichia coli*. *Pharmacol.Biochem.Behav.* **13**, 191-195

Jones, R.P. (1989). Biological principles for the effects of ethanol. *Enzyme.Microb.Tech.* **11**, 130-153

Jordan, D.B., Bowman, M.J., Braker, J.D., Dien, B.S., Hector, R.E., Lee, C.C., Mertens, J.A., & Wagschal, K. (2012). Plant cell walls to ethanol. *Biochem.J.* **442**, 241-252

Kai, Y., Kashiwagi, T., Ishikawa, K., Ziyatdinov, M.K., Redkina, E.I., Kiriukhin, M.Y., Gusyatiner, M.M., Kobayashi, S., Takagi, H., & Suzuki, E. (2006). Engineering of *Escherichia coli* L-serine O-acetyltransferase on the basis of crystal structure: desensitization to feedback inhibition by L-cysteine. *Protein Eng Des Sel* **19**, 163-167

Kapust, R.B. & Waugh, D.S. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8**, 1668-1674

Kazahaya, T., Kawai, K., Yashima, S., & Sasaki, Y. (1972). Aerobic dissimilation of glucose by heterolactic bacteria III. aldehyde dehydrogenase and alcohol dehydrogenase of *Leuconostoc mesenteroides*. *J.Gen.Appl.Microbiol.* **18**, 43-55

Kessler, D., Herth, W., & Knappe, J. (1992). Ultrastructure and pyruvate formate-lyase radical quenching property of the multienzymic AdhE protein of *Escherichia coli*. *J.Biol.Chem.* **267**, 18073-18079

Kessler, D., Leibrecht, I., & Knappe, J. (1991). Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase activities of *Escherichia coli* reside on a polymeric protein particle encoded by adhE. *FEBS Lett.* **281**, 59-63

Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L., & Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* **37**, D387-D392

Koo, O.K., Jeong, D.W., Lee, J.M., Kim, M.J., Lee, J.H., Chang, H.C., Kim, J.H., & Lee, H.J. (2005). Cloning and characterization of the bifunctional alcohol/acetaldehyde dehydrogenase gene (*adhE*) in *Leuconostoc mesenteroides* isolated from kimchi. *Biotechnol.Lett.* **27**, 505-510

Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J.Mol.Biol.* **372**, 774-797

Langendorf, C.G., Key, T.L., Fenalti, G., Kan, W.T., Buckle, A.M., Caradoc-Davies, T., Tuck, K.L., Law, R.H., & Whisstock, J.C. (2010). The X-ray crystal structure of *Escherichia coli* succinic semialdehyde dehydrogenase; structural insights into $NADP^+$/enzyme interactions. *PLoS One* **5**, e9280

Langer, G., Cohen, S.X., Lamzin, V.S., & Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat.Protoc.* **3**, 1171-1179

Lei, Y., Pawelek, P.D., & Powlowski, J. (2008). A shared binding site for NAD^+ and coenzyme A in an acetaldehyde dehydrogenase involved in bacterial degradation of aromatic compounds. *Biochemistry* **47**, 6870-6882

Leonardo, M.R., Dailly, Y., & Clark, D.P. (1996). Role of NAD in regulating the *adhE* gene of *Escherichia coli*. *J.Bacteriol.* **178**, 6013-6018

Limayem, A. & Ricke, S.C. (2012). Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. *Prog.Energ.Combust.* **38**, 449-467

Liu, S. & Qureshi, N. (2009). How microbes tolerate ethanol and butanol. *N.Biotechnol.* **26**, 117-121

Long, F., Vagin, A.A., Young, P., & Murshudov, G.N. (2008). BALBES: a molecular-replacement pipeline. *Acta Crystallogr.D.Biol.Crystallogr.* **64**, 125-132

Luque, R., Herrero-Davila, L., Campelo, J.M., Clark, J.H., Hidalgo, J.M., Luna, D., Marinas, J.M., & Romero, A.A. (2008). Biofuels: a technological perspective. *Energy Environ.Sci.* **1**, 542-564

Lynd, L.R., Laser, M.S., Bransby, D., Dale, B.E., Davison, B., Hamilton, R., Himmel, M., Keller, M., McMillan, J.D., Sheehan, J., & Wyman, C.E. (2008). How biotech can transform biofuels. *Nat.Biotechnol.* **26**, 169-172

Malloy, A. & Carr, B. (2008). Nanoparticle measurement through visualisation. *Seps.Science.*

Martin, M.A. (2010). First generation biofuels compete. *N.Biotechnol.* **27**, 596-608

Matayoshi, S., Oda, H., & Sarwar, G. (1989). Relationship between the production of spirosomes and anaerobic glycolysis activity in *Escherichia coli* B. *J.Gen.Microbiol.* **135**, 525-529

Meher, L.C., Vidya Sagar, D., & Naik, S.N. (2006). Technical aspects of biodiesel production by transesterification-a review. *Renew.Sust.Energ.Rev.* **10**, 248-268

Membrillo-Hernandez, J., Echave, P., Cabiscol, E., Tamarit, J., Ros, J., & Lin, E.C. (2000). Evolution of the *adhE* gene product of *Escherichia coli* from a functional reductase to a dehydrogenase. Genetic and biochemical studies of the mutant proteins. *J.Biol.Chem.* **275**, 33869-33875

Membrillo-Hernandez, J. & Lin, E.C. (1999). Regulation of expression of the *adhE* gene, encoding ethanol oxidoreductase in *Escherichia coli*: transcription from a downstream promoter and regulation by fnr and RpoS. *J.Bacteriol.* **181**, 7571-7579

MiTeGen. MiTeGen's MicroRT™ capillary system. http://www.mitegen.com/products/micrort/micrort.shtml

Nagi, A.D. & Regan, L. (1997). An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold.Des.* **2**, 67-75

Nair, R.V., Bennett, G.N., & Papoutsakis, E.T. (1994). Molecular characterization of an aldehyde/alcohol dehydrogenase gene from *Clostridium acetobutylicum* ATCC 824. *J.Bacteriol.* **176**, 871-885

Nave, C. (1995). Radiation damage in protein crystallography. *Radiat.Phys.Chem.* **45**, 483-490

Neves, A.R., Ramos, A., Costa, H., van Swam, I.I., Hugenholtz, J., Kleerebezem, M., de, V.W., & Santos, H. (2002). Effect of different NADH oxidase levels on glucose metabolism by *Lactococcus lactis*: kinetics of intracellular metabolite pools determined by in vivo nuclear magnetic resonance. *Appl.Environ.Microbiol.* **68**, 6332-6342

Nicholas, K.B., Nicholas, H.B.J., & Deerfield, D.W.I. (1997). GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNEW.NEWS* **4**,

Nnyepi, M.R., Peng, Y., & Broderick, J.B. (2007). Inactivation of *E. coli* pyruvate formate-lyase: role of AdhE and small molecules. *Arch.Biochem.Biophys.* **459**, 1-9

Notredame, C., Higgins, D.G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J.Mol.Biol.* **302**, 205-217

Novagen. pET System Manual. 11th. 2011.

Novy, R. (2001). Overcoming the codon bias of *E.coli* for enhanced protein expression. *Innovations*

Paget, M. S. 2012. University of Sussex, Sussex House, Brighton, BN1 9RH. Unpublished Data quoted with permission.

Pei, J., Zhou, Q., Jiang, Y., Le, Y., Li, H., Shao, W., & Wiegel, J. (2010). *Thermoanaerobacter* spp. control ethanol pathway via transcriptional regulation and versatility of key enzymes. *Metab.Eng.* **12**, 420-428

Pei, J., Zhou, Q., Jing, Q., Li, L., Dai, C., Li, H., Wiegel, J., & Shao, W. (2011). The mechanism for regulating ethanol fermentation by redox levels in *Thermoanaerobacter ethanolicus*. *Metab.Eng.* **13**, 186-193

Peng, H., Wu, G., & Shao, W. (2008). The aldehyde/alcohol dehydrogenase (AdhE) in relation to the ethanol formation in *Thermoanaerobacter ethanolicus* JW200. *Anaerobe.* **14**, 125-127

Perham, R.N. (1975). Self-assembly of biological macromolecules. *Philos.Trans.R.Soc.Lond B Biol.Sci.* **272**, 123-136

Perozich, J., Nicholas, H., Wang, B.C., Lindahl, R., & Hempel, J. (1999). Relationships within the aldehyde dehydrogenase extended family. *Protein.Sci.* **8**, 137-146

Potocnik, J. (2007). Renewable energy sources and the realities of setting an energy agenda. *Science.* **315**, 810-811

Potterton, E., Briggs, P., Turkenburg, M., & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr.D.Biol.Crystallogr.* **59**, 1131-1137

Promega. Wizard Plus SV Minipreps DNA Purification System. 2010a.

Promega. Wizard SV Gel and PCR Clean-Up System. 2010b.

RCSB. (25-4-2012). PDB. http://www.rcsb.org/pdb/ligand/ligandsummary.do?hetId=CSO&sid=2RFV

Richter, N., Zienert, A., & Hummel, W. (2011). A single-point mutation enables lactate dehydrogenase from *Bacillus subtilis* to utilize NAD^+ and $NADP^+$ as cofactor. *Eng.Life Sci.* **11**, 26-36

Ritchie, D.W. & Kemp, G.J.L. (1999). Fast Computation, Rotation, and Comparison of Low Resolution Spherical Harmonic Molecular Surfaces. *J.Comp.Chem.* **20**, 383-395

Rossmann, M.G., Moras, D., & Olsen, K.W. (1974). Chemical and biological evolution of nucleotide-binding protein. *Nature.* **250**, 194-199

Russell, R.J., Ferguson, J.M., Hough, D.W., Danson, M.J., & Taylor, G.L. (1997). The crystal structure of citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9 Å resolution. *Biochemistry.* **36**, 9983-9994

Sali, A. & Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J.Mol.Biol.* **234**, 779-815

Sanchez, L.B. (1998). Aldehyde dehydrogenase (CoA-acetylating) and the mechanism of ethanol formation in the amitochondriate protist, *Giardia lamblia*. *Arch.Biochem.Biophys.* **354**, 57-64

Sanchez, O.J. & Cardona, C.A. (2008). Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresour.Technol.* **99**, 5270-5295

Sawers, G. & Watson, G. (1998). A glycyl radical solution: oxygen-dependent interconversion of pyruvate formate-lyase. *Mol.Microbiol.* **29**, 945-954

Schoffelen, S. & van Hest, J.C.M. (2011). Multi-enzyme systems: bringing enzymes together in vitro. *Soft Matter.* 1736-1746

Schwarzenbacher, R., von, D.F., Canaves, J.M., Brinen, L.S., Dai, X., Deacon, A.M., Elsliger, M.A., Eshaghi, S., Floyd, R., Godzik, A., Grittini, C., Grzechnik, S.K., Guda, C., Jaroszewski, L., Karlak, C., Klock, H.E., Koesema, E., Kovarik, J.S., Kreusch, A., Kuhn, P., Lesley, S.A., McMullan, D., McPhillips, T.M., Miller, M.A., Miller, M.D., Morse, A., Moy, K., Ouyang, J., Page, R., Robb, A., Rodrigues, K., Selby, T.L., Spraggon, G., Stevens, R.C., van den Bedem, H., Velasquez, J., Vincent, J., Wang, X., West, B., Wolf, G., Hodgson, K.O., Wooley, J., & Wilson, I.A. (2004). Crystal structure of an iron-containing 1,3-propanediol dehydrogenase (TM0920) from *Thermotoga maritima* at 1.3 Å resolution. *Proteins.* **54**, 174-177

Searchinger, T., Heimlich, R., Houghton, R.A., Dong, F., Elobeid, A., Fabiosa, J., Tokgoz, S., Hayes, D., & Yu, T.H. (2008). Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science.* **319**, 1238-1240

Shao, X., Raman, B., Zhu, M., Mielenz, J.R., Brown, S.D., Guss, A.M., & Lynd, L.R. (2011). Mutant selection and phenotypic and genetic characterization of ethanol-tolerant strains of *Clostridium thermocellum*. *Appl.Microbiol.Biotechnol.* **92**, 641-652

Shaw, A.J., Jenney, F.E., Adams, M.W.W., & Lynd, L.R. (2008). End-product pathways in the xylose fermenting bacterium, *Thermoanaerobacterium saccharolyticum*. *Enzyme.Microb.Tech.* **42**, 453-458

Shevchenko, A., Wilm, M., Vorm, O., & Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal.Chem.* **68**, 850-858

Shone, C.C. & Fromm, H.J. (1981). Steady-state and pre-steady-state kinetics of coenzyme A linked aldehyde dehydrogenase from *Escherichia coli*. *Biochemistry.* **20**, 7494-7501

Sickmier, E.A., Brekasis, D., Paranawithana, S., Bonanno, J.B., Paget, M.S., Burley, S.K., & Kielkopf, C.L. (2005). X-ray structure of a Rex-family repressor/NADH complex insights into the mechanism of redox sensing. *Structure.* **13**, 43-54

Sigrist, C.J., Cerutti, L., de, C.E., Langendijk-Genevaux, P.S., Bulliard, V., Bairoch, A., & Hulo, N. (2010). PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids.Res.* **38**, D161-D166

Smith, L.T. & Kaplan, N.O. (1980). Purification, properties, and kinetic mechanism of coenzyme A-linked aldehyde dehydrogenase from *Clostridium kluyveri*. *Arch.Biochem.Biophys.* **203**, 663-675

Sporty, J.L., Kabir, M.M., Turteltaub, K.W., Ognibene, T., Lin, S.J., & Bench, G. (2008). Single sample extraction protocol for the quantification of NAD^+ and NADH redox states in *Saccharomyces cerevisiae*. *J.Sep.Sci.* **31**, 3202-3211

Srere, P.A. (1975). The enzymology of the formation and breakdown of citrate. *Adv.Enzymol.Relat Areas Mol.Biol.* **43**, 57-101

Stratagene. ArcticExpress™ RIL Competent Cells and ArcticExpress™ RP Competent Cells Instruction Manual. 2009.

Taylor, M.P., Eley, K.L., Martin, S., Tuffin, M.I., Burton, S.G., & Cowan, D.A. (2009). Thermophilic ethanologenesis: future prospects for second-generation bioethanol production. *Trends Biotechnol.* **27**, 398-405

Taylor, M.P., Esteban, C.D., & Leak, D.J. (2008). Development of a versatile shuttle vector for gene expression in *Geobacillus* spp. *Plasmid.* **60**, 45-52

The World Bank Group. (14-5-2012). World Bank Data Website. http://data.worldbank.org/

Timmons, M.D., Knutson, B.L., Nokes, S.E., Strobel, H.J., & Lynn, B.C. (2009). Analysis of composition and structure of *Clostridium thermocellum* membranes from wild-type and ethanol-adapted strains. *Appl.Microbiol.Biotechnol.* **82**, 929-939

TMO Renewables. 1-4-2009, 10-1-2012 & 1-7-2012. 40 Alan Turing Road, Surrey Research Park, Guildford. GU2 7YF. Personal Communications

Tomita, T., Kuzuyama, T., & Nishiyama, M. (2006). Alteration of coenzyme specificity of lactate dehydrogenase from *Thermus thermophilus* by introducing the loop region of NADP(H)-dependent malate dehydrogenase. *Biosci.Biotechnol.Biochem.* **70**, 2230-2235

Vieille, C. & Zeikus, G.J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol.Mol.Biol.Rev.* **65**, 1-43

Wang, E., Bauer, M.C., Rogstam, A., Linse, S., Logan, D.T., & von, W.C. (2008). Structure and functional properties of the *Bacillus subtilis* transcriptional repressor Rex. *Mol.Microbiol.* **69**, 466-478

Wharton, C. & Eisenthal, R. 1981. *Molecular Enzymology* Wiley.

Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. *J.Appl.Crystallogr.* **43**, 186-190

Wolfe, A.J. (2005). The acetate switch. *Microbiol.Mol.Biol.Rev.* **69**, 12-50

Zhang, Y.H. (2011). Substrate channeling and enzyme complexes for biotechnological applications. *Biotechnol.Adv.* **29**, 715-725

Gene and corresponding amino acid sequences

G. thermoglucosidasius ADHE

DNA coding sequence:

ATGCGTATGGCTGTGGAGGAGAGAGTCGTCGATAAAAAAATCGAAGTAGCAAAAATGATTGATGAGCTTGTCGCTA
ATGCACAGAAAGCGTTGGAACAAATTCGCGCTTACGATCAAGAAACGATCGATCATATCGTGAAAGAAATGGCGTT
AGCCGGGCTCGACAAGCATATGGCATTAGCCAAGCTTGCAGTAGAAGAAACAAAACGCGGTGTATATGAAGATAAA
ATCATAAAAAACCTTTTTGCGACAGAATATATATACCACAATATTAAGTATGATAAAACAGTCGGGATTATTCATGAAA
ATCCGCATGAAGAAATTATCGAAATTGCTGAGCCTGTTGGTGTTATTGCTGGGATTACGCCAGTGACAAACCCGAC
ATCGACAACGATGTTTAAAGCGTTAATCTCGATAAAAACACGCAACCCGATTATTTTCGCTTTCCATCCATCGGCGC
AACGATGCAGCAGCGAAGCGGCAAGAGTGCTGCGCGATGCGGCGGTCCGGGCAGGGGCTCCAGAACATTGCATT
CAATGGATTGAAACTCCTTCGCTTGATGCAACCAATCAGCTTATGCACCATCCTGGCGTTTCTCTCATTTTGGCAAC
TGGTGGCGCCGGCATGGTGAAAGCAGCGTACAGCTCTGGAAAACCAGCTTTGGGCGTCGGACCTGGCAATGTGC
CTTGCTATATTGAAAAAACGGCAAACATAAAACGGGCGGTAAATGACTTAATTTTATCGAAAACGTTTGATAACGGC
ATGATTTGCGCTTCTGAACAAGCAGTCATTATTGATAAAGAAATTTATGAACAAGTAAAGAAAGAAATGATAGAAAAC
CATTGTTATTTCTTAAATGAAGAAGAAAAGAAAAAAGTAGAAAAACTCGTTATCAATGAAAATACATGCGCCGTCAAC
CCGGATATCGTCGGAAAGCCAGCTTATGAAATTGCGAAAATGGCCGGCATCGCTGTGCCGGAAGACACAAAAATTC
TTGTTGCTGAGTTAAAAGGGGTCGGGCCAAAATATCCGTTGTCTCGGGAAAAATTAAGCCCTGTCCTTGCTTGCTAT
AAAGTTAACAGCACGGAAGAAGGATTTAAGCGCTGTGAAGAAATGCTGGAATTTGGCGGCTTGGGACATTCGGCTG
TCATCCATTCCGATAATCAAAACGTGGTTACCGAATTTGGCAAACGGATGAAAGCGGGACGGATTATCGTTAATGC
GCCATCTTCGCAAGGAGCAATCGGCGATATTTACAATGCGTACATTCCGTCATTAACGCTGGGATGCGGCACATTT
GGCGGAAACTCTGTTTCGACAAACGTCAGTGCGATTCATCTTATCAATATAAAAAGAATGGCAAAAAGGACGGTAAA
TATGCAATGGTTTAAAGTGCCGCCGAAAATTTATTTCGAAAAAAATGCTGTACAATACTTAGCGAAAATGCCGGATAT
TTCCAGAGCTTTTATCGTCACCGACCCGGGAATGGTCAAGCTCGGATATGTCGATAAAGTGCTGTATTACTTGCGC
AGACGCCCGGATTATGTGCATAGTGAAATTTTCTCCGAAGTAGAGCCAGATCCTTCAATTGAGACGGTAATGAAAG
GTGTCGATATGATGAGAAGTTTCGAGCCGGATGTGATTATCGCGCTTGGAGGCGGCTCGCCAATGGATGCGGCAA
AAGCGATGTGGCTCTTTTACGAGCATCCGACAGCGGATTTCAACGCATTAAAACAAAAATTTTTAGATATTCGAAAA
CGCGTTTATAAATATCCAAAACTGGGCCAAAAAGCGAAATTTGTCGCCATTCCGACGACATCAGGAACAGGATCGG
AAGTAACGTCCTTTGCCGTCATTACCGATAAAAAAACGAATATAAAATATCCGTTGGCAGATTATGAATTGACACCG
GACGTCGCGATTGTGGATCCGCAATTTGTCATGACCGTGCCAAAACATGTCACCGCCGATACGGGAATGGATGTAT
TGACACATGCGATCGAAGCGTATGTCTCCAATATGGCAAATGATTATACCGATGGTCTTGCCATGAAAGCAATCCAA
CTCGTATTTGAATATTTGCCGCGGGCATATCAAAACGGAGCGGATGAGCTTGCCCGGGAGAAAATGCATAACGCCT
CTACGATTGCGGGAATGGCATTTGCCAACGCGTTTTTAGGCATTAACCATAGTTTGGCTCATAAACTTGGCGCGGA
ATTCCATATTCCGCATGGGCGCGCGAATACCATTTTGATGCCGCATGTCATTCGCTATAACGCAGCGAAACCGAAA
AAATTTACCGCATTTCCGAAATACGAATATTTCAAAGCGGACCAGCGCTATGCAGAAATTGCGAGAATGCTCGGCTT
GCCGGCCCGCACAACGGAAGAAGGGGTCGAAAGCCTCGTTCAGGCGATCATTAAGCTGGCAAAACAGTTGGATAT
GCCGCTGAGCATTGAAGCATGCGGCGTCAGCAAACAAGAATTTGAAAGCAAAGTTGAAAAATTAGCCGAATTGGCT
TTCGAAGACCAATGTACTACTGCTAACCCGAAACTCCCGCTTGTTAGCGATTTAGTTCATATTTATCGCCAAGCGTT
TAAAGGAGTTTAA

Amino acid sequence:

MRMAVEERVVDKKIEVAKMIDELVANAQKALEQIRAYDQETIDHIVKEMALAGLDKHMALAKLAVEETKRGVYEDKIIKNL
FATEYIYHNIKYDKTVGIIHENPHEEIIEIAEPVGVIAGITPVTNPTSTTMFKALISIKTRNPIIFAFHPSAQRCSSEAARVLRDA
AVRAGAPEHCIQWIETPSLDATNQLMHHPGVSLILATGGAGMVKAAYSSGKPALGVGPGNVPCYIEKTANIKRAVNDLIL
SKTFDNGMICASEQAVIIDKEIYEQVKKEMIENHCYFLNEEEKKKVEKLVINENTCAVNPDIVGKPAYEIAKMAGIAVPEDT
KILVAELKGVGPKYPLSREKLSPVLACYKVNSTEEGFKRCEEMLEFGGLGHSAVIHSDNQNVVTEFGKRMKAGRIIVNAP
SSQGAIGDIYNAYIPSLTLGCGTFGGNSVSTNVSAIHLINIKRMAKRTVNMQWFKVPPKIYFEKNAVQYLAKMPDISRAFIV
TDPGMVKLGYVDKVLYYLRRRPDYVHSEIFSEVEPDPSIETVMKGVDMMRSFEPDVIIALGGGSPMDAAKAMWLFYEHP
TADFNALKQKFLDIRKRVYKYPKLGQKAKFVAIPTTSGTGSEVTSFAVITDKKTNIKYPLADYELTPDVAIVDPQFVMTVPK
HVTADTGMDVLTHAIEAYVSNMANDYTDGLAMKAIQLVFEYLPRAYQNGADELAREKMHNASTIAGMAFANAFLGINHSL
AHKLGAEFHIPHGRANTILMPHVIRYNAAKPKKFTAFPKYEYFKADQRYAEIARMLGLPARTTEEGVESLVQAIIKLAKQLD
MPLSIEACGVSKQEFESKVEKLAELAFEDQCTTANPKLPLVSDLVHIYRQAFKGV-

Fragment 11

DNA coding sequence:

ATGAATATGCAATGGTTTAAAGTGCCGCCGAAAATTTATTTCGAAAAAAATGCTGTACAATACTTAGCGAAAATGCC
GGATATTTCCAGAGCTTTTATCGTCACCGACCCGGGAATGGTCAAGCTCGGATATGTCGATAAAGTGCTGTATTACT
TGCGCAGACGCCCGGATTATGTGCATAGTGAAATTTTCTCCGAAGTAGAGCCAGATCCTTCAATTGAGACGGTAAT
GAAAGGTGTCGATATGATGAGAAGTTTCGAGCCGGATGTGATTATCGCGCTTGGAGGCGGCTCGCCAATGGATGC
GGCAAAAGCGATGTGGCTCTTTTACGAGCATCCGACAGCGGATTTCAACGCATTAAAACAAAAATTTTTAGATATTC
GAAAACGCGTTTATAAATATCCAAAACTGGGCCAAAAAGCGAAATTTGTCGCCATTCCGACGACATCAGGAACAGG
ATCGGAAGTAACGTCCTTTGCCGTCATTACCGATAAAAAAACGAATATAAAATATCCGTTGGCAGATTATGAATTGA
CACCGGACGTCGCGATTGTGGATCCGCAATTTGTCATGACCGTGCCAAAACATGTCACCGCCGATACGGGAATGG

ATGTATTGACACATGCGATCGAAGCGTATGTCTCCAATATGGCAAATGATTATACCGATGGTCTTGCCATGAAAGCA
ATCCAACTCGTATTTGAATATTTGCCGCGGGCATATCAAAACGGAGCGGATGAGCTTGCCCGGGAGAAAATGCATA
ACGCCTCTACGATTGCGGGAATGGCATTTGCCAACGCGTTTTTAGGCATTAACCATAGTTTGGCTCATAAACTTGGC
GCGGAATTCCATATTCCGCATGGGCGCGCGAATACCATTTTGATGCCGCATGTCATTCGCTATAACGCAGCGAAAC
CGAAAAAATTTACCGCATTTCCGAAATACGAATATTTCAAAGCGGACCAGCGCTATGCAGAAATTGCGAGAATGCTC
GGCTTGCCGGCCCGCACAACGGAAGAAGGGGTCGAAAGCCTCGTTCAGGCGATCATTAAGCTGGCAAAACAGTTG
GATATGCCGCTGAGCATTGAAGCATGCGGCGTCAGCAAACAAGAATTTGAAAGCAAAGTTGAAAAATTAGCCGAAT
TGGCTTTCGAAGACCAATGTACTACTGCTAACCCGAAACTCCCGCTTGTTAGCGATTTAGTTCATATTTATCGCCAA
GCGTTTAAAGGAGTTTAA

Amino acid sequence:

MNMQWFKVPPKIYFEKNAVQYLAKMPDISRAFIVTDPGMVKLGYVDKVLYYLRRRPDYVHSEIFSEVEPDPSIETVMKGV
DMMRSFEPDVIIALGGGSPMDAAKAMWLFYEHPTADFNALKQKFLDIRKRVYKYPKLGQKAKFVAIPTTSGTGSEVTSFA
VITDKKTNIKYPLADYELTPDVAIVDPQFVMTVPKHVTADTGMDVLTHAIEAYVSNMANDYTDGLAMKAIQLVFEYLPRAY
QNGADELAREKMHNASTIAGMAFANAFLGINHSLAHKLGAEFHIPHGRANTILMPHVIRYNAAKPKKFTAFPKYEYFKAD
QRYAEIARMLGLPARTTEEGVESLVQAIIKLAKQLDMPLSIEACGVSKQEFESKVEKLAELAFEDQCTTANPKLPLVSDLV
HIYRQAFKGV-

EutE

DNA coding sequence:

ATGAGCGTGGATGCACAAAAAATTGAGAAACTTGTAAGAAAAATACTGGAGGAAATGGAAGAGAAAAAGAAGCCGG
CAGAAACCGAATGCGAATGGGGCATATTTGACCATATGAATCAAGCAATTGAAGCAGCTGAAATCGCCCAAAAAGA
ACTTGTGCAACTATCTCTTGGGCAAAGAGGGAAACTGATTGAGGCAATTCGCAAAGCTGCCAAGGAGAATGCGGAA
AAATTTGCAAGAATGGCAGTGGACGAAACAGGGATGGGTAAATATGAAGATAAGATAGTAAAAAATTTACTTGCAGC
CGAAAAAACTCCAGGCATTGAAGACTTGCGGACAGAAGTTTTTTCTGGCGATGATGGTTTAACGCTTGTTGAACTTT
CGCCTTATGGTGTTATAGGAGCGATCACACCGACAACAAATCCTACTGAGACGATTATTTGTAATTCAATCGGAATG
ATCGCTGCAGGAAATGCCGTTGTTTTCAGCCCGCATCCGAGAGCAAAAAATACATCTTTGTATGCGATTAAAATTTT
CAATCAAGCAATCGTTGAAGCAGGCGGGCCGAAGAACTTGATAACAACGGTTGCAAATCCTTCCATTGAGCAAGCG
GAGATCATGATGAAACATAAAACCATTAAAATGCTGGTGGCTACCGGAGGGCCAGGAGTGGTAAAAGCAGTGCTCT
CAAGCGGGAAAAAAGCGATAGGTGCCGGTGCTGGAAATCCGCCAGTAGTTGTGGATGAGACGGCAGATATTGAAA
AGGCAGCGAAAGATATTATCGCTGGATGCAGCTTTGACAATAATCTTCCGTGTGTTGCAGAAAAAGAAGTAATTGCT
GTCGAATCGATTGCCGACCGTTTAATTGATTATATGAAAAAACACGGTGCATATGAAATAACGAATAAAGAACAAATA
CAACAACTAACTGATTTAGTGGTAGAGAATGGACATGCCAATAAAGAATTTGTTGGTAAAGATGCAGCCTATATTTTA
AAACATATTGGCATCAATGTTCCGCCTGACATCCGTGTAGCAATAATGGAAGTAGATGGTAAACATCCATTAGTAAC
AGTCGAGTTAATGATGCCAATTTTGCCTATTGTCCGCGTCAAAAACGTGGATCAAGCTATCGAGCTAGCTGTGGAA
GTTGAACATGGGTTCCGACATACGGCGATTATGCATTCTAAAAATGTAGATCATTTAACAAAATTCGCAAAGGCAAT
CCAAACTACGATTTTTGTAAAAAATGCTCCATCTTATGCCGGCATTGGCGTAGGCGGTGAAGGATATGCTACCTTCA
CCATAGCGGGGCCTACAGGTGAAGGCTTAACTTCAGCAAAAGACTTTGCAAGAAAAAGAAAATGTGTCTTAGTGGA
TGCTTTGTCGATAAGATAA

Amino acid sequence:

MSVDAQKIEKLVRKILEEMEEKKKPAETECEWGIFDHMNQAIEAAEIAQKELVQLSLGQRGKLIEAIRKAAKENAEKFARM
AVDETGMGKYEDKIVKNLLAAEKTPGIEDLRTEVFSGDDGLTLVELSPYGVIGAITPTTNPTETIICNSIGMIAAGNAVVFSP
HPRAKNTSLYAIKIFNQAIVEAGGPKNLITTVANPSIEQAEIMMKHKTIKMLVATGGPGVVKAVLSSGKKAIGAGAGNPPVV
VDETADIEKAAKDIIAGCSFDNNLPCVAEKEVIAVESIADRLIDYMKKHGAYEITNKEQIQQLTDLVVENGHANKEFVGKDA
AYILKHIGINVPPDIRVAIMEVDGKHPLVTVELMMPILPIVRVKNVDQAIELAVEVEHGFRHTAIMHSKNVDHLTKFAKAIQT
TIFVKNAPSYAGIGVGGEGYATFTIAGPTGEGLTSAKDFARKRKCVLVDALSIR-

AcAldDH

DNA coding sequence:

ATGTTGCGTGACATCGATTTGCAATCCATCCAAGAGGTCAGAAATTATCTTGAAGAAGCGAAAGCAGCACAAAAAAT
CCTTGAAAAAATGACACAAAGCGAGATTGACAAGATTGTCGAAAGCATGGCAAATGCAGCGAGAGAAGAGGCTGG
CCGTTTAGCAGCGATGGCAGTCGAAGAGACCGGCTTCGGGAATGTGGAAGATAAAACGTTGAAAAATTTGTTCGCA
GCGAATGATGTTTACAACTCTATCAAAGATGTGAAAACAGTTGGGATTATTCGCCGTGATGAAGAAAACCGCGTTTG
GGAAATTGCTCAACCTGTCGGGATTGTTGCAGGAATCATCCCATCTACCAACCCAACTTCAACAGTTATTTTTAAAG
CATTGATCGCAGTTAAAGCAAGAAATGCCATTGTTTTCAGTCCGCACCCATCAGCAGCAAAATGTACAGCAGAAGC
AGCAAGAATCATGCAAGAAGCAGCGGAGCGCGCCGGAGCGCCAAAAGGGTTGATTTCTTGCATCACACAACCGAC
AATGGCTGCAACAAACGAGTTAATGAAACATAAATTAACTGACGTCATTTTAGCAACAGGCGGCCCTGGTTTGGTGA
AAGCAGCGTACAGTTCCGGAAAACCAGCATATGGCGTTGGGCCTGGAAACGTGCCAGTATATATCCATGAAAGTGC
CAATATCGCCAAAGCTGTTCAATTAATCATCCAAAGCAAGACTTTTGACTATGGAACCATTTGCGCTTCTGAACAAG
CTCTTTTAGTGGATGAGTCGATTAAAGAAAAAGTGGTTGCTGAGTTAAAACAACAAGGCGCTTACTTCTTGAATGAA
GAAGAAAAACAAAAAGTAGCATCCATCATTATGGTTAACGGTTCATTAAATGCGAAGATCGTTGGAAAAGCGCCGCA
AGTAATCGCGGAAATGGCTGGGATTGAGATTCCATCTGACGTGAAGTTGCTTGTGGCAGAAGAAACAGAAGTGGG
GAAAGAATATCCATTCTCCATTGAAAAATTGTCTCCAATTCTAGCATTCTATATAGTGAAAGGGATGGAAGAAGCAA
GCGAGCTTGCCCAAAAATTGCTTGAAGTGGGCGGACTCGGCCATACAGTTGGAATTCACGCTGAAGATGAAAAAGT
GATCGAAGCGTACACCATCGATAAACCAGCCGGACGGATTGTTGTAAATGCTGGTACAACGTTTGGCGGAATTGGT
GCAACAGTGAATGTTAAACCATCCTTGACACTTGGATGCGGTGCTATCGGTAACAACATTACATCAGACAACGTGAC
AGTAACTCATTTATTCAACATTAAACGCGTGGCATTTGGAGTCCGCGAGATGCCAAAAAAAGTTGAAGGCGCACAAA
AAGAACCTGCATTAACTAAATAA

Amino acid sequence:

MLRDIDLQSIQEVRNYLEEAKAAQKILEKMTQSEIDKIVESMANAAREEAGRLAAMAVEETGFGNVEDKTLKNLFAANDV
YNSIKDVKTVGIIRRDEENRVWEIAQPVGIVAGIIPSTNPTSTVIFKALIAVKARNAIVFSPHPSAAKCTAEAARIMQEAAERA
GAPKGLISCITQPTMAATNELMKHKLTDVILATGGPGLVKAAYSSGKPAYGVGPGNVPVYIHESANIAKAVQLIIQSKTFDY
GTICASEQALLVDESIKEKVVAELKQQGAYFLNEEEKQKVASIIMVNGSLNAKIVGKAPQVIAEMAGIEIPSDVKLLVAEETE
VGKEYPFSIEKLSPILAFYIVKGMEEASELAQKLLEVGGLGHTVGIHAEDEKVIEAYTIDKPAGRIVVNAGTTFGGIGATVNV
KPSLTLGCGAIGNNITSDNVTVTHLFNIKRVAFGVREMPKKVEGAQKEPALTK-

Fusion 1 (SacI site removed)

DNA coding sequence: (AcAldDH sequence underlined)

ATGAATATGCAATGGTTTAAAGTGCCGCCGAAAATTTATTTCGAAAAAAATGCTGTACAATACTTAGCGAAAATGCC
GGATATTTCCAGAGCTTTTATCGTCACCGACCCGGGAATGGTCAAGCTCGGATATGTCGATAAAGTGCTGTATTACT
TGCGCAGACGCCCGGATTATGTGCATAGTGAAATTTTCTCCGAAGTAGAGCCAGATCCTTCAATTGAGACGGTAAT
GAAAGGTGTCGATATGATGAGAAGTTTCGAGCCGGATGTGATTATCGCGCTTGGAGGCGGCTCGCCAATGGATGC
GGCAAAAGCGATGTGGCTCTTTTACGAGCATCCGACAGCGGATTTCAACGCATTAAAACAAAAATTTTTAGATATTC
GAAAACGCGTTTATAAATATCCAAAACTGGGCCAAAAAGCGAAATTTGTCGCCATTCCGACGACATCAGGAACAGG
ATCGGAAGTAACGTCCTTTGCCGTCATTACCGATAAAAAAACGAATATAAAATATCCGTTGGCAGATTATGAATTGA
CACCGGACGTCGCGATTGTGGATCCGCAATTTGTCATGACCGTGCCAAAACATGTCACCGCCGATACGGGAATGG
ATGTATTGACACATGCGATCGAAGCGTATGTCTCCAATATGGCAAATGATTATACCGATGGTCTTGCCATGAAAGCA
ATCCAACTCGTATTTGAATATTTGCCGCGGGCATATCAAAACGGAGCGGATGAGCTTGCCCGGGAGAAAATGCATA
ACGCCTCTACGATTGCGGGAATGGCATTTGCCAACGCGTTTTTAGGCATTAACCATAGTTTGGCTCATAAACTTGGC
GCGGAATTCCATATTCCGCATGGGCGCGCGAATACCATTTTGATGCCGCATGTCATTCGCTATAACGCAGCGAAAC
CGAAAAAATTTACCGCATTTCCGAAATACGAATATTTCAAAGCGGACCAGCGCTATGCAGAAATTGCGAGAATGCTC
GGCTTGCCGGCCCGCACAACGGAAGAAGGGGTCGAAAGCCTCGTTCAGGCGATCATTAAGCTGGCAAAACAGTTG
GATATGCCGCTGAGCATTGAAGCATGCGGCGTCAGCAAACAAGAATTTGAAAGCAAAGTTGAAAAATTAGCCGAAT
TGGCTTTCGAAGACCAATGTACTACTGCTAACCCGAAACTCCCGCTTGTTAGCGATTTAGTTCATATTTATCGCCAA
GCGTTTAAAGGAGTTCGGAGCGGTTCCAACATGTTGCGTGACATCGATTTGCAATCCATCCAAGAGGTCAGAAATT
ATCTTGAAGAAGCGAAAGCAGCACAAAAAATCCTTGAAAAAATGACACAAAGCGAGATTGACAAGATTGTCGAAAG
CATGGCAAATGCAGCGAGAGAAGAGGCTGGCCGTTTAGCAGCGATGGCAGTCGAAGAGACCGGCTTCGGGAATG
TGGAAGATAAAACGTTGAAAAATTTGTTCGCAGCGAATGATGTTTACAACTCTATCAAAGATGTGAAAACAGTTGGG
ATTATTCGCCGTGATGAAGAAAACCGCGTTTGGGAAATTGCTCAACCTGTCGGGATTGTTGCAGGAATCATCCCAT
CTACCAACCCAACTTCAACAGTTATTTTTAAAGCATTGATCGCAGTTAAAGCAAGAAATGCCATTGTTTTCAGTCCGC
ACCCATCAGCAGCAAAATGTACAGCAGAAGCAGCAAGAATCATGCAAGAAGCAGCGGAGCGCGCCGGAGCGCCA
AAAGGGTTGATTTCTTGCATCACACAACCGACAATGGCTGCAACAAACGAGTTAATGAAACATAAATTAACTGACGT
CATTTTAGCAACAGGCGGCCCTGGTTTGGTGAAAGCAGCGTACAGTTCCGGAAAACCAGCATATGGCGTTGGGCC
TGGAAACGTGCCAGTATATATCCATGAAAGTGCCAATATCGCCAAAGCTGTTCAATTAATCATCCAAAGCAAGACTT
TTGACTATGGAACCATTTGCGCTTCTGAACAAGCTCTTTTAGTGGATGAGTCGATTAAAGAAAAAGTGGTTGCTGAG
TTAAAACAACAAGGCGCTTACTTCTTGAATGAAGAAGAAAAACAAAAAGTAGCATCCATCATTATGGTTAACGGTTCA
TTAAATGCGAAGATCGTTGGAAAAGCGCCGCAAGTAATCGCGGAAATGGCTGGGATTGAGATTCCATCTGACGTGA
AGTTGCTTGTGGCAGAAGAAACAGAAGTGGGGAAAGAATATCCATTCTCCATTGAAAAATTGTCTCCAATTCTAGCA
TTCTATATAGTGAAAGGGATGGAAGAAGCAAGCGAGCTTGCCCAAAAATTGCTTGAAGTGGGCGGACTCGGCCATA
CAGTTGGAATTCACGCTGAAGATGAAAAAGTGATCGAAGCGTACACCATCGATAAACCAGCCGGACGGATTGTTGT
AAATGCTGGTACAACGTTTGGCGGAATTGGTGCAACAGTGAATGTTAAACCATCCTTGACACTTGGATGCGGTGCT
ATCGGTAACAACATTACATCAGACAACGTGACAGTAACTCATTTATTCAACATTAAACGCGTGGCATTTGGAGTCCG
CGAGATGCCAAAAAAAGTTGAAGGCGCACAAAAAGAACCTGCATTAACTAAATAA

Amino acid sequence: (AcAldDH sequence underlined)

MNMQWFKVPPKIYFEKNAVQYLAKMPDISRAFIVTDPGMVKLGYVDKVLYYLRRRPDYVHSEIFSEVEPDPSIETVMKGV
DMMRSFEPDVIIALGGGSPMDAAKAMWLFYEHPTADFNALKQKFLDIRKRVYKYPKLGQKAKFVAIPTTSGTGSEVTSFA
VITDKKTNIKYPLADYELTPDVAIVDPQFVMTVPKHVTADTGMDVLTHAIEAYVSNMANDYTDGLAMKAIQLVFEYLPRAY
QNGADELAREKMHNASTIAGMAFANAFLGINHSLAHKLGAEFHIPHGRANTILMPHVIRYNAAKPKKFTAFPKYEYFKAD
QRYAEIARMLGLPARTTEEGVESLVQAIIKLAKQLDMPLSIEACGVSKQEFESKVEKLAELAFEDQCTTANPKLPLVSDLV
HIYRQAFKGVRSGSNMLRDIDLQSIQEVRNYLEEAKAAQKILEKMTQSEIDKIVESMANAAREEAGRLAAMAVEETGFGN
VEDKTLKNLFAANDVYNSIKDVKTVGIIRRDEENRVWEIAQPVGIVAGIIPSTNPTSTVIFKALIAVKARNAIVFSPHPSAAKC
TAEAARIMQEAAERAGAPKGLISCITQPTMAATNELMKHKLTDVILATGGPGLVKAAYSSGKPAYGVGPGNVPVYIHESA
NIAKAVQLIIQSKTFDYGTICASEQALLVDESIKEKVVAELKQQGAYFLNEEEKQKVASIIMVNGSLNAKIVGKAPQVIAEMA
GIEIPSDVKLLVAEETEVGKEYPFSIEKLSPILAFYIVKGMEEASELAQKLLEVGGLGHTVGIHAEDEKVIEAYTIDKPAGRIV
VNAGTTFGGIGATVNVKPSLTLGCGAIGNNITSDNVTVTHLFNIKRVAFGVREMPKKVEGAQKEPALTK-

Fusion 2

DNA coding sequence: (AcAldDH sequence underlined)

ATGTTGCGTGACATCGATTTGCAATCCATCCAAGAGGTCAGAAATTATCTTGAAGAAGCGAAAGCAGCACAAAAAAT
CCTTGAAAAAATGACACAAAGCGAGATTGACAAGATTGTCGAAAGCATGGCAAATGCAGCGAGAGAAGAGGCTGG
CCGTTTAGCAGCGATGGCAGTCGAAGAGACCGGCTTCGGGAATGTGGAAGATAAAACGTTGAAAAATTTGTTCGCA
GCGAATGATGTTTACAACTCTATCAAAGATGTGAAAACAGTTGGGATTATTCGCCGTGATGAAGAAAACCGCGTTTG
GGAAATTGCTCAACCTGTCGGGATTGTTGCAGGAATCATCCCATCTACCAACCCAACTTCAACAGTTATTTTTAAAG
CATTGATCGCAGTTAAAGCAAGAAATGCCATTGTTTTCAGTCCGCACCCATCAGCAGCAAAATGTACAGCAGAAGC

AGCAAGAATCATGCAAGAAGCAGCGGAGCGCGCCGGAGCGCCAAAAGGGTTGATTTCTTGCATCACACAACCGAC
AATGGCTGCAACAAACGAGTTAATGAAACATAAATTAACTGACGTCATTTTAGCAACAGGCGGCCCTGGTTTGGTGA
AAGCAGCGTACAGTTCCGGAAAACCAGCATATGGCGTTGGGCCTGGAAACGTGCCAGTATATATCCATGAAAGTGC
CAATATCGCCAAAGCTGTTCAATTAATCATCCAAAGCAAGACTTTTGACTATGGAACCATTTGCGCTTCTGAACAAG
CTCTTTTAGTGGATGAGTCGATTAAAGAAAAAGTGGTTGCTGAGTTAAAACAACAAGGCGCTTACTTCTTGAATGAA
GAAGAAAAACAAAAAGTAGCATCCATCATTATGGTTAACGGTTCATTAAATGCGAAGATCGTTGGAAAAGCGCCGCA
AGTAATCGCGGAAATGGCTGGGATTGAGATTCCATCTGACGTGAAGTTGCTTGTGGCAGAAGAAACAGAAGTGGG
GAAAGAATATCCATTCTCCATTGAAAAATTGTCTCCAATTCTAGCATTCTATATAGTGAAAGGGATGGAAGAAGCAA
GCGAGCTTGCCCAAAAATTGCTTGAAGTGGGCGGACTCGGCCATACAGTTGGAATTCACGCTGAAGATGAAAAAGT
GATCGAAGCGTACACCATCGATAAACCAGCCGGACGGATTGTTGTAAATGCTGGTACAACGTTTGGCGGAATTGGT
GCAACAGTGAATGTTAAACCATCCTTGACACTTGGATGCGGTGCTATCGGTAACAACATTACATCAGACAACGTGAC
AGTAACTCATTTATTCAACATTAAACGCGTGGCATTTGGAGTCCGCGAGATGCCAAAAAAAGTTGAAGGCGCACAAA
AAGAACCTGCATTAACTAAAAAGCTTATGAATATGCAATGGTTTAAAGTGCCGCCGAAAATTTATTTCGAAAAAAATG
CTGTACAATACTTAGCGAAAATGCCGGATATTTCCAGAGCTTTTATCGTCACCGACCCGGGAATGGTCAAGCTCGG
ATATGTCGATAAAGTGCTGTATTACTTGCGCAGACGCCCGGATTATGTGCATAGTGAAATTTTCTCCGAAGTAGAGC
CAGATCCTTCAATTGAGACGGTAATGAAAGGTGTCGATATGATGAGAAGTTTCGAGCCGGATGTGATTATCGCGCT
TGGAGGCGGCTCGCCAATGGATGCGGCAAAAGCGATGTGGCTCTTTTACGAGCATCCGACAGCGGATTTCAACGC
ATTAAAACAAAAATTTTTAGATATTCGAAAACGCGTTTATAAATATCCAAAACTGGGCCAAAAAGCGAAATTTGTCGC
CATTCCGACGACATCAGGAACAGGATCGGAAGTAACGTCCTTTGCCGTCATTACCGATAAAAAAACGAATATAAAAT
ATCCGTTGGCAGATTATGAATTGACACCGGACGTCGCGATTGTGGATCCGCAATTTGTCATGACCGTGCCAAAACA
TGTCACCGCCGATACGGGAATGGATGTATTGACACATGCGATCGAAGCGTATGTCTCCAATATGGCAAATGATTAT
ACCGATGGTCTTGCCATGAAAGCAATCCAACTCGTATTTGAATATTTGCCGCGGGCATATCAAAACGGAGCGGATG
AGCTTGCCCGGGAGAAAATGCATAACGCCTCTACGATTGCGGGAATGGCATTTGCCAACGCGTTTTTAGGCATTAA
CCATAGTTTGGCTCATAAACTTGGCGCGGAATTCCATATTCCGCATGGGCGCGCGAATACCATTTTGATGCCGCAT
GTCATTCGCTATAACGCAGCGAAACCGAAAAAATTTACCGCATTTCCGAAATACGAATATTTCAAAGCGGACCAGCG
CTATGCAGAAATTGCGAGAATGCTCGGCTTGCCGGCCCGCACAACGGAAGAAGGGGTCGAAAGCCTCGTTCAGGC
GATCATTAAGCTGGCAAAACAGTTGGATATGCCGCTGAGCATTGAAGCATGCGGCGTCAGCAAACAAGAATTTGAA
AGCAAAGTTGAAAAATTAGCCGAATTGGCTTTCGAAGACCAATGTACTACTGCTAACCCGAAACTCCCGCTTGTTAG
CGATTTAGTTCATATTTATCGCCAAGCGTTTAAAGGAGTTTAA

Amino acid sequence: (AcAldDH sequence underlined)

MLRDIDLQSIQEVRNYLEEAKAAQKILEKMTQSEIDKIVESMANAAREEAGRLAAMAVEETGFGNVEDKTLKNLFAANDV
YNSIKDVKTVGIIRRDEENRVWEIAQPVGIVAGIIPSTNPTSTVIFKALIAVKARNAIVFSPHPSAAKCTAEAARIMQEAAERA
GAPKGLISCITQPTMAATNELMKHKLTDVILATGGPGLVKAAYSSGKPAYGVGPGNVPVYIHESANIAKAVQLIIQSKTFDY
GTICASEQALLVDESIKEKVVAELKQQGAYFLNEEEKQKVASIIMVNGSLNAKIVGKAPQVIAEMAGIEIPSDVKLLVAEETE
VGKEYPFSIEKLSPILAFYIVKGMEEASELAQKLLEVGGLGHTVGIHAEDEKVIEAYTIDKPAGRIVVNAGTTFGGIGATVNV
KPSLTLGCGAIGNNITSDNVTVTHLFNIKRVAFGVREMPKKVEGAQKEPALTKKLMNMQWFKVPPKIYFEKNAVQYLAKM
PDISRAFIVTDPGMVKLGYVDKVLYYLRRRPDYVHSEIFSEVEPDPSIETVMKGVDMMRSFEPDVIIALGGGSPMDAAKA
MWLFYEHPTADFNALKQKFLDIRKRVYKYPKLGQKAKFVAIPTTSGTGSEVTSFAVITDKKTNIKYPLADYELTPDVAIVDP
QFVMTVPKHVTADTGMDVLTHAIEAYVSNMANDYTDGLAMKAIQLVFEYLPRAYQNGADELAREKMHNASTIAGMAFAN
AFLGINHSLAHKLGAEFHIPHGRANTILMPHVIRYNAAKPKKFTAFPKYEYFKADQRYAEIARMLGLPARTTEEGVESLVQ
AIIKLAKQLDMPLSIEACGVSKQEFESKVEKLAELAFEDQCTTANPKLPLVSDLVHIYRQAFKGV-

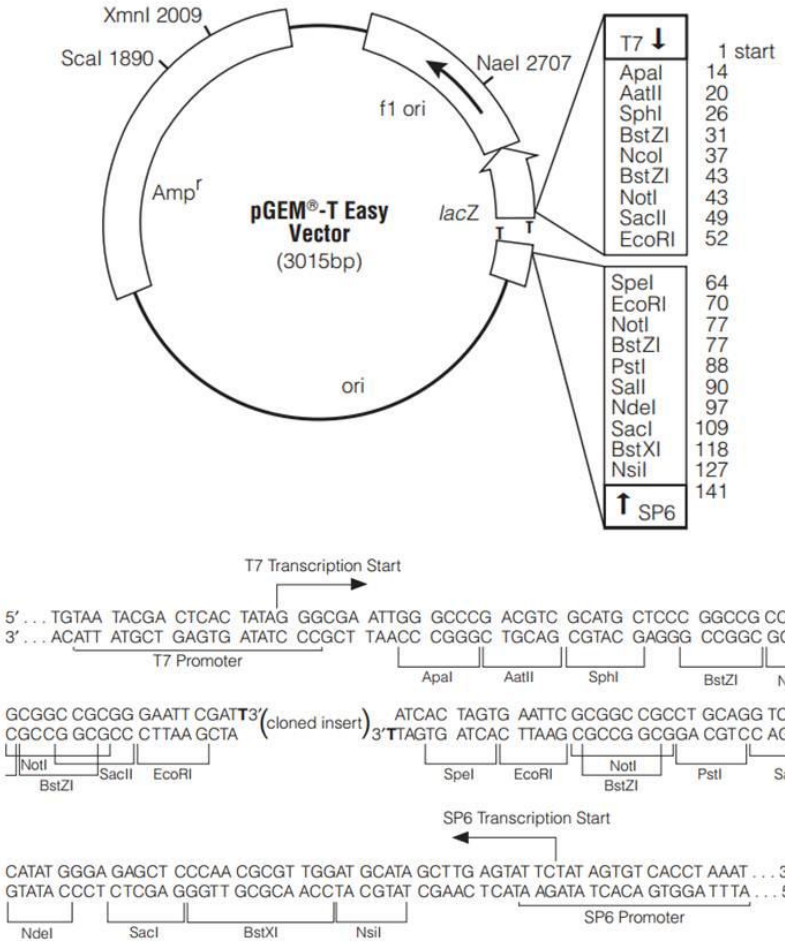
DNA primer details

Primer name	Restriction site	DNA Sequence (5’-3’)
C-Tag ADHE F1	NcoI	CATCCGCCATGGTGCGTATGGCTGTGGAGGAGAG AGTCG
C-Tag ADHE R1	XhoI	CATCCGCTCGAGAACTCCTTTAAACGCTTGG
ADHE F1 MUT	n/a	CTTTAAGAAGGAGATATACCATGCGTATGGCTGTG GAGGAGAG
ADHE F1R MUT	n/a	CTCTCCTCCACAGCCATACGCATGGTATATCTCCTT CTTAAAG
Fragment 1 (AldDH) F	NheI	CTAGCTAGCATGCGTATGGCTGTGGAGG
Fragment 1 (AldDH) R	XhoI	CCGCTCGAGCTACGTCTCAATTGAAGGATCTGG
Fragment 2 (ADH) F	NdeI	CCGCATATGATTTACAATGCGTACATTCCG
Fragment 2 (ADH) R	XhoI	CCGCTCGAGTTAAACTCCTTTAAACGCTTGG
Fragment 3 (AldDHs) F	NheI	CTAGCTAGCATGCGTATGGCTGTGGAGG
Fragment 3 (AldDHs) R	XhoI	CCGCTCGAGTTAATGAATCGCACTGACGTTTGTCG
Frag11pET28aF	NdeI	CCGCATATGATTTACATGAATATGCAATGGTTTAAA G
Frag11pET28aR	XhoI	CCCTCGAGGCTTAAACTCCTTTAAACGCT
pUCG18F	n/a	CAATAGTCAAACAATCGCCACAA
pUCG18Rev	n/a	TATGCTTCCGGCTCGTAT
GB ADHE Fwd1	XbaI	GCTCTAGAGCATGCGTATGGCTGTGG
GB ADHE Rev1	SacI	CGAGCTCGTTAAACTCCTTTAAACGCT

GB ADH Fwd1 (Frag 2)	XbaI	GCTCTAGAGCATGATTTACAATGCGTACATTCCG
GB aldDHshort Rev1 (Frag 3)	SacI	CGAGCTCGTTAATGAATCGCACTGACG
GB 1-345aa Rev1 (Frag 4)	SacI	CGAGCTCGTTACAACGGATATTTTGGC
GB 1-400aa Rev1 (Frag 5)	SacI	CGAGCTCGTTATTTCATCCGTTTGCC
GB 1-505aa Rev1 (Frag 6)	SacI	CGAGCTCGTTACACTTTATCGACATATC
GB 1-600aa Rev1 (Frag 7)	SacI	CGAGCTCGTTAAAATTTCGCTTTTTGGC
GB 1-745aa Rev1 (Frag 8)	SacI	CGAGCTCGTTACCCATGCGGAATATG
GB 1-800aa Rev1 (Frag 9)	SacI	CGAGCTCGTTATTCGACCCCTTCTTCC
GB 517-869aa Fwd1 (Frag 10)	XbaI	GCTCTAGAGCATGAGTGAAATTTTCTCCG
GB 459-869aa Fwd1 (Frag 11)	XbaI	GCTCTAGAGCATGAATATGCAATGGTTTAAAG
GB 267-869aa Fwd1 (Frag 12)	XbaI	GCTCTAGAGCATGGAAATTTATGAACAAGTAAAG
eutEpET45bF	KpnI	GGGGTACCATGAGCGTGGATGCACAA
eutEpET45bR	XhoI	CCCTCGAGGGTTATCTTATCGACAAAGCATCCACTA
eutEpUCG18F	XbaI	CTCTAGAGCATGAGCGTGGATGCACAAAAAATTG
eutEpUCG18R	SacI	CGAGCTCGTTATCTTATCGACAAAGCATCCACTA
eutE pLM303	KpnI	GGGTACCCATGAGCGTGGATGCACAA
eutE pLM303 (seq)	n/a	CAGTTTCCCTCTTTGCCCAAG
GB ACaldDH Fwd1	XbaI	GCTCTAGAGATGTTGCGTGACATCGATTTGC
GB ACaldDH Rev1	SacI	CGAGCTCGTTATTTAGTTAATGCAGGTTCTTTTTG
ACaldDH pet28 F1	NheI	GCTAGCATGTTGCGTGACATCGATTTGC
ACaldDH pet28 rev1	XhoI	CCGCTCGAGTTATTTAGTTAATGCAGGTTCTTTTTG
Frag11 for eutE fusion F1	NcoI	CCATGGCATGAATATGCAATGGTTTAAAG
Frag11 for eutE fusion R1	SacI	GAGCTCCGAACTCCTTTAAACGCTTG
ACaldDHfusion Fwd1 (pLM303)	SacI	GAGCTCCAACATGTTGCGTGACATCG
11-acaldDH Fus int1	n/a	CGGAAGAAGGGGTCGAA
11-acaldDH Fus int2rev	n/a	CATCATTCGCTGCG
pUC F1 Rev1	SacI	CGAGCTCGTTATTTAGTTAATGCAGGTTCTTTTTG
11-AcaldDH NO SACI Fwd	n/a	TTAAAGGAGTTCGGAGCGGTTCCAACATGTTGCGTGA
11-AcaldDH NO SACI Rev	n/a	TCACGCAACATGTTGGAACCGCTCCGAACTCCTTTAA
Frag 11 28a Fusion 2 F1	HindIII	AAGCTTATGAATATGCAATGGTTTAAAG
AcAldDH 28a Fusion2 R1	HindIII	AAGCTTTTTAGTTAATGCAGGTTCTTTT
pUC F2 Fwd1	XbaI	TCTAGAATGTTGCGTGACATCGATTTGC
pUC F2 Rev1	SacI	CGAGCTCGTTAAACTCCTTTAAACGCT
Fus2 int1	n/a	GTTGGAAAAGCGCC
Fus2 mut 1 F	n/a	CAAAAAGAACCTGCATTAACTAAAAAGGGTCTTATGAATATGCAATGGTTTAAAGTG

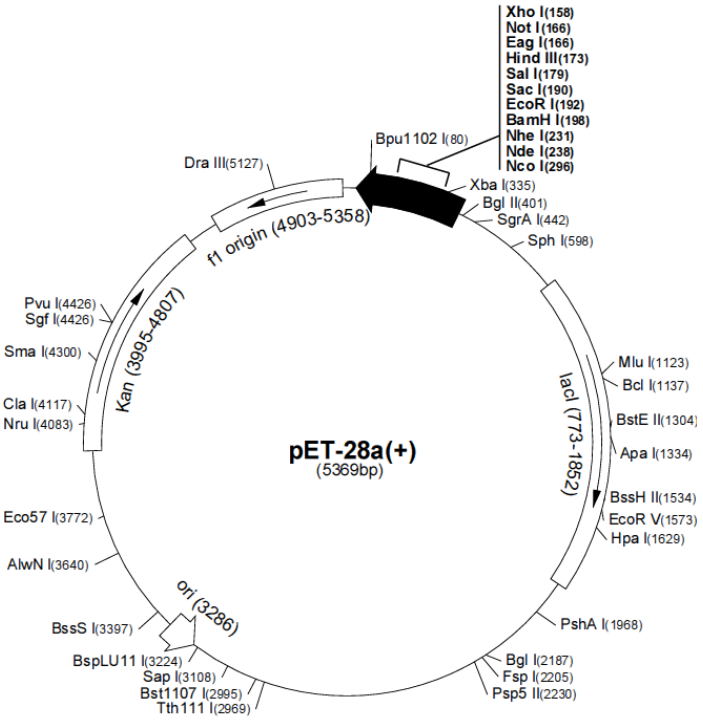
Fus2 mut 1 R	n/a	CACTTTAAACCATTGCATATTCATAAGACCCTTTTTA GTTAATGCAGGTTCTTTTTG
AcAldDH pET28a R1 Native link	HindIII	AAGCTTTGCCACGCGTTTAATGTTGA
ADH native link F1	HindIII	AAGCTTACGGTAAATATGCAATGGTTTA
Fus4 Fwd primer	n/a	ACATTAAACGCGTGGCAAAGGGGAGCCTTACGGTA AATATGCAATG
Fus4 Rev primer	n/a	CATTGCATATTTACCGTAAGGCTCCCCTTTGCCACG CGTTTAATGT
AcAldDH-11 int rev	n/a	TCGATCGCATGTGTCA
Fus4 int F 2	n/a	GCAAGAAATGCCATTG
pADHEfwd1	pstI	CTGCAGTAGCTTAAATGAGGTTTTAGAC
pADHErev1	XbaI	TCTAGATTCTCCCTCCTGATTGTGAA
pUCG18 prom F1	n/a	ATCGGTGCGGGCCT

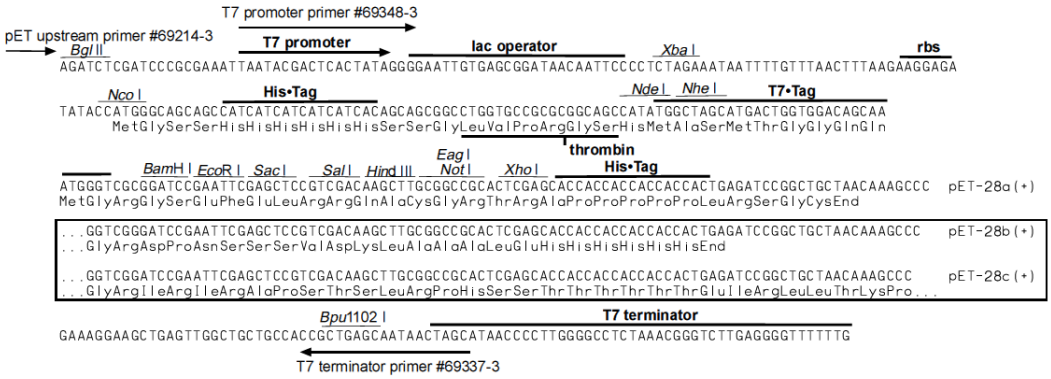
Vector maps:



pGEM®-T easy promoter/multiple cloning site sequence

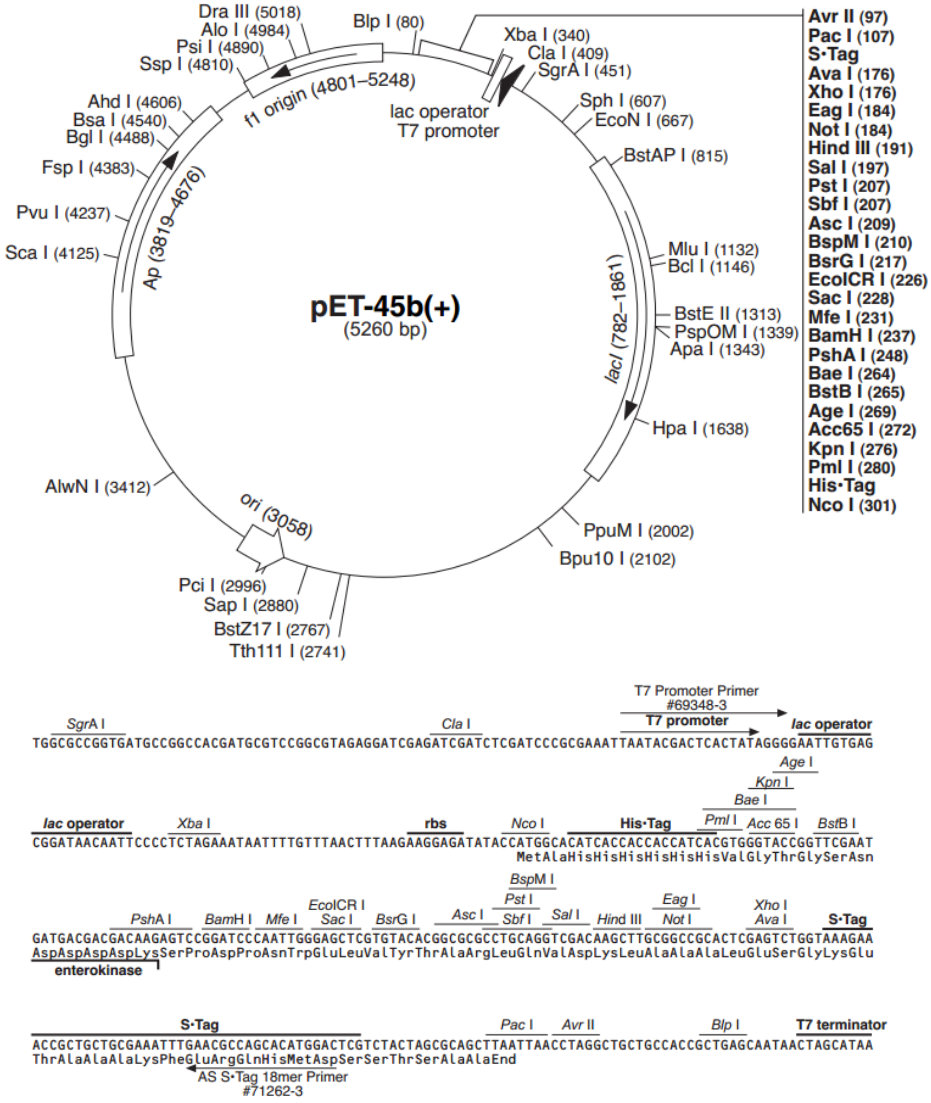
pGEM®-T easy vector map and multiple cloning site details. Image adapted from product handbook (Promega).





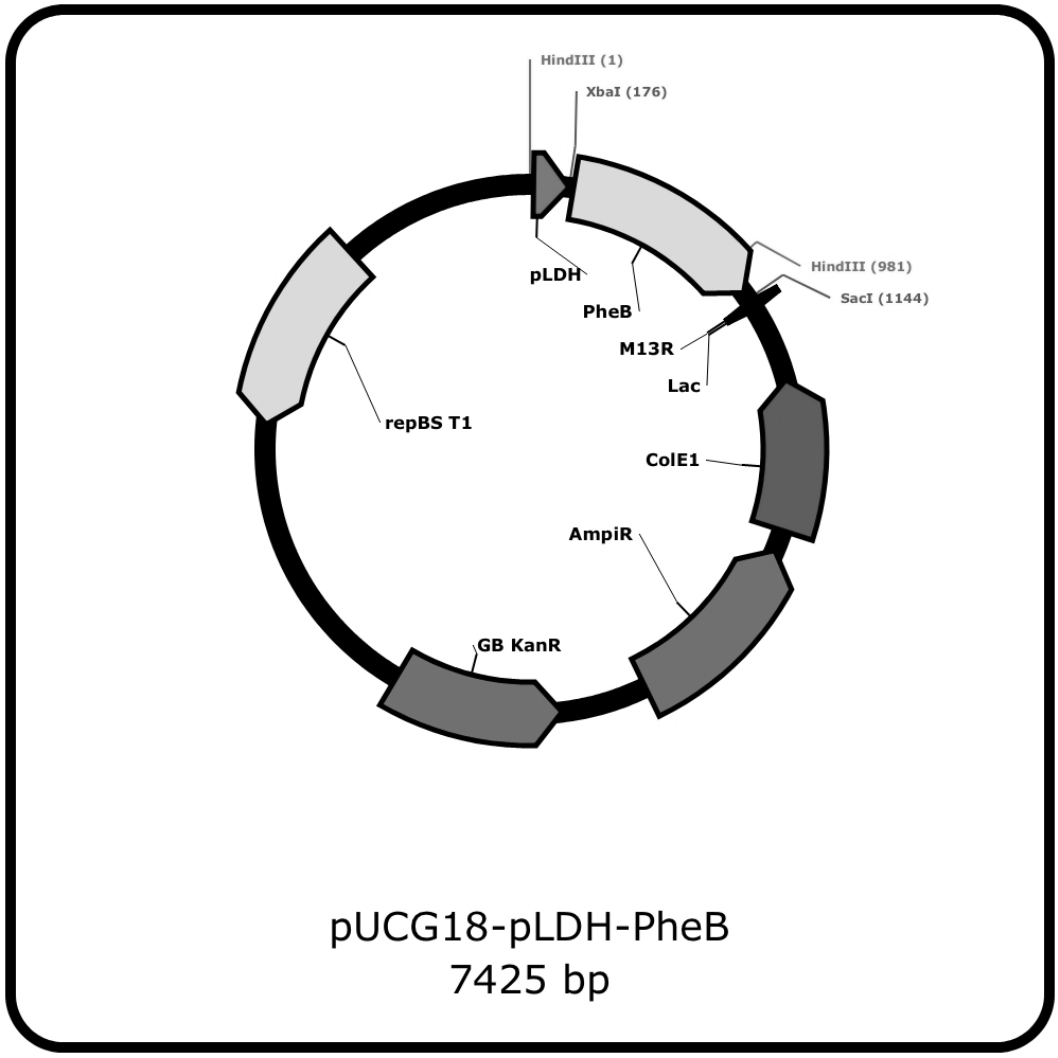
pET-28a-c(+) cloning/expression region

pET28a vector map and multiple cloning site details. Image adapted from product handbook (Novagen).

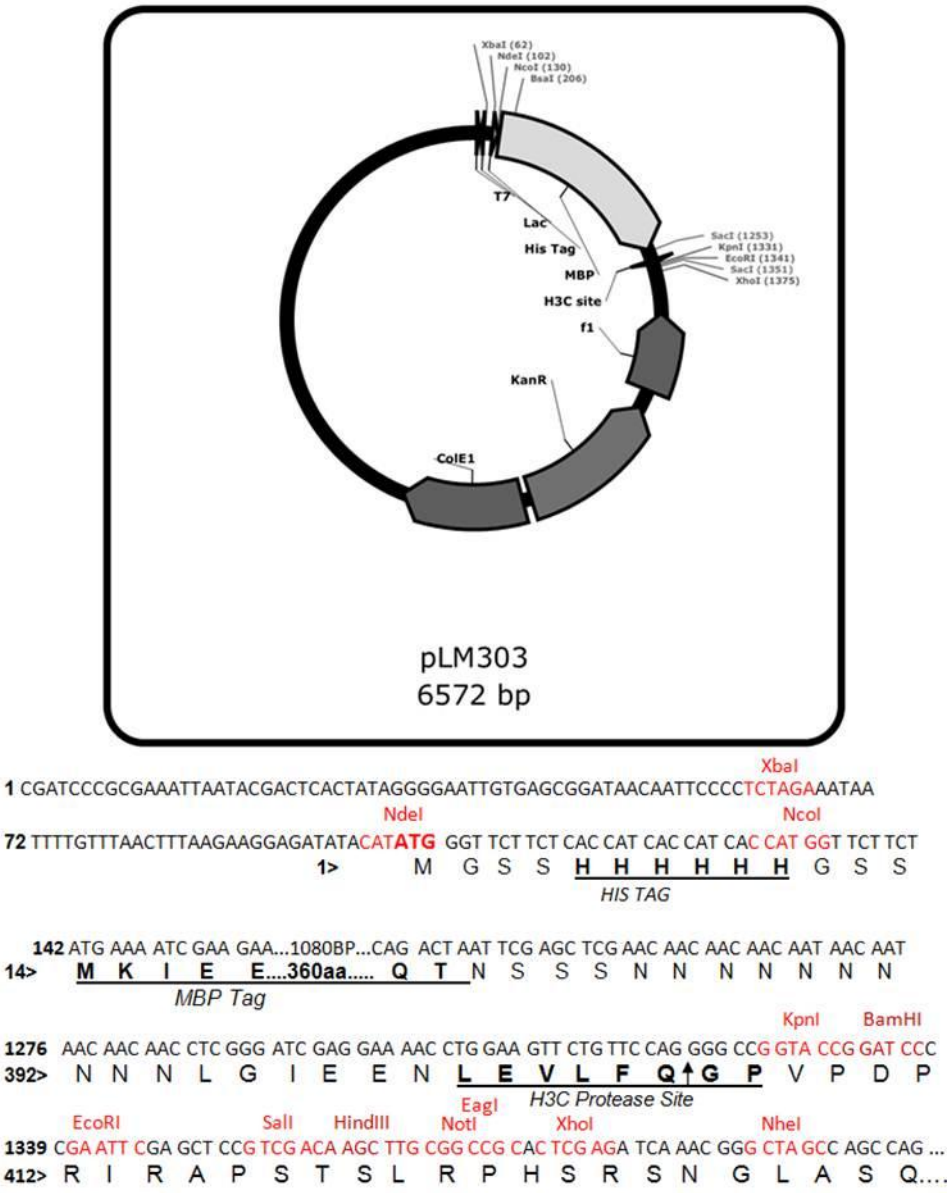


pET-45b(+) cloning/expression region

pET45b vector map and multiple cloning site details. Image adapted from product handbook (Novagen).



pUCG18-pLDH-*pheB* vector map. Image produced from the vector sequence using the Plasma DNA programme (Angers-Loustau et al. 2007).



XbaI

1 CGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAA

NdeI NcoI

72 TTTTGTTTAACTTTAAGAAGGAGATATACATATG GGT TCT TCT CAC CAT CAC CAT CAC CAT GGT TCT TCT

1> M G S S H H H H H H G S S

HIS TAG

142 ATG AAA ATC GAA GAA...1080BP...CAG ACT AAT TCG AGC TCG AAC AAC AAC AAC AAT AAC AAT

14> M K I E E....360aa..... Q T N S S S N N N N N N N

MBP Tag

KpnI BamHI

1276 AAC AAC AAC CTC GGG ATC GAG GAA AAC CTG GAA GTT CTG TTC CAG GGG CCG GTA CCG GAT CCC

392> N N N L G I E E N L E V L F Q↑G P V P D P

H3C Protease Site

EagI

EcoRI SalI HindIII NotI XhoI NheI

1339 CGA ATT CGA GCT CCG TCG ACA AGC TTG CGG CCG CAC TCG AGA TCA AAC GGG CTA GCC AGC CAG ...

412> R I R A P S T S L R P H S R S N G L A S Q....

pLM303 vector map and multiple cloning site details. Image produced from the vector sequence using the Plasma DNA programme (Angers-Loustau et al. 2007).